RANDOM DISTRIBUTION FUNCTIONS

LESTER E. DUBINS and DAVID A. FREEDMAN
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction and summary

How can one choose, at random, a probability measure on the unit interval? This paper develops the answer announced in [4]. Section 1, which can be skipped without logical loss, gives a fairly full but slightly informal account. The formalities begin with section 2. All later sections are largely independent of one another. Section 10 indexes definitions made in one section but used in other sections.

A *distribution function* F on the closed unit interval I is a nondecreasing, nonnegative, real-valued function on I, normalized to be 1 at 1 and continuous from the right. To each F there corresponds one and only one probability measure $|F|$ on the Borel subsets of I, with $F(x)$ equal to the $|F|$-measure of the closed interval $[0, x]$, for all $x \in I$. Choosing at random a probability on I is therefore tantamount to choosing at random a distribution function on I.

A *random distribution function* $\mathbf{F}$ is a measurable map from a probability space $(\Omega, \mathfrak{F}, Q)$ to the space Δ of distribution functions on the closed unit interval I, where Δ is endowed with its natural Borel σ-field, that is, the σ-field generated by the customary weak* topology. The distribution of $\mathbf{F}$, namely $Q\mathbf{F}^{-1}$, is a *prior* probability measure on Δ. Of course, $\mathbf{F}$ is essentially the stochastic process $\{F_t, 0 \leq t \leq 1\}$ on $(\Omega, \mathfrak{F}, Q)$, whose sample functions are distribution functions: $F_t(\omega)$ is $\mathbf{F}(\omega)$ evaluated at t. Therefore, this paper can be thought of as dealing with a class of random distribution functions, with a class of stochastic processes, or with a class of prior probabilities. Similar priors are treated in [9], [11], [16], and [17].

Since the indefinite integral of a distribution function is convex, this paper can also be thought of as dealing with a class of random convex functions, but we do not pursue this idea.

Which class of random distribution functions does this paper study? A *base probability* μ is a probability on the Borel subsets of the unit square S, assigning measure 0 to the corners $(0, 0)$ and $(1, 1)$. Each such μ determines a random distribution function $\mathbf{F}$ and so a prior probability P_μ on Δ, which will now be described, by explaining how to select a value of $\mathbf{F}$, that is, a distribution function F, at random.

ASSUMPTION. *For ease of exposition, we assume throughout this section that μ concentrates on, that is, assigns probability 1 to, the interior of S.*

This paper was supported in part by the National Science Foundation Grant GP-2593, and in part by a grant to Freedman from the Sloan Foundation.

Construction. To select a value F of $\mathbf{F}$ at random, begin by selecting a point (x, y) from the interior of S according to μ. The horizontal and vertical lines through (x, y) divide S into four rectangles. Consider the closed lower left rectangle L and the upper right one R. The unique (*positive, affine*) transformation of the form $(u, v) \to (\alpha u + \beta, \gamma v + \delta)$, α and γ positive, which maps S onto L, carries μ into a probability μ_L concentrated on L. The probability μ_R is defined similarly. Now select a point (x_L, y_L) at random from the interior of L according to μ_L, and a point (x_R, y_R) at random from the interior of R according to μ_R. As before, (x_L, y_L) determines four subrectangles of L, and (x_R, y_R) determines four subrectangles of R. Consider the lower left subrectangle LL of L, the upper right subrectangle RL of L, and the analogous subrectangles LR and RR of R. The construction may be continued by selecting one point at random from each of these four rectangles, according to the appropriate affine image of μ, and so on. This procedure yields a nested decreasing sequence of closed sets, the n-th one being a union of 2^n closed rectangles, namely, S, $L \cup R$, $LL \cup RL \cup LR \cup RR$, and so on. The intersection of these closed sets is a nonempty closed set which, with probability 1, is the graph of a distribution function. This function is taken as the random value F of $\mathbf{F}$.

Section 2 gives a formal description of this construction, including a proof that the closed set in question really is the graph of a distribution function. The idea of the proof is to show that the set has area 0 with probability 1, because the sets shrinking to it have areas whose expectations shrink to 0.

We do not know an abstract characterization of the set of priors P_μ as μ ranges over the base probabilities.

Section 3 gives a necessary and sufficient condition on μ for P_μ to assign positive mass to every nonempty open subset of Δ. When the support of μ (the smallest closed set of μ-probability 1) contains neither $(0, 0)$ nor $(1, 1)$, the result is easy to state: the support of P_μ is then all of Δ if and only if the graph of every $F \in \Delta$ meets the support of μ. We conjecture that unless the support of P_μ is all of Δ, it has empty interior.

As shown in section 4, P_μ assigns probability 1 to the continuous and strictly increasing distribution functions. For at stage n of the construction, there are 2^n closed rectangles whose union includes the graph of the distribution function being constructed. The sum of the squares of the heights (respectively, widths) of these 2^n rectangles decreases, has expectation converging to 0, and therefore converges to 0 almost everywhere. Consequently, so does the maximum height (respectively, width).

Unless μ concentrates on the main diagonal of S, P_μ concentrates on the purely singular distribution functions. For many μ, Kinney and Pitcher [14] have sharpened this result beautifully (by showing that P_μ-almost all F have a Hausdorff dimension which is a constant less than 1). Here is a different sharpening (section 5). Say $F \in \Delta$ is *strictly singular* if F has a finite, positive derivative nowhere. Unless μ concentrates on the main diagonal, P_μ-almost all F are strictly singular. The case which we learned from de Rham of a μ that assigns measure 1 to a single off-diagonal point, say (r, w), brings out one of the ideas in the proof. Clearly, P_μ then assigns measure 1 to a single distribution function, say $S_{w,r}$. (This two-parameter family of distribution functions is studied in ([8], chapters 5 and 6), and the one-parameter family $S_{w,1/2}$ is studied in [2].) Fix $x \in I$. At stage n of the construction of $S_{w,r}$, there will be 2^n closed rectangles whose union includes the graph of $S_{w,r}$. Of these rectangles, there will be a leftmost one whose projection on the horizontal axis contains x. Its diagonal d_n is a chord in the graph of $S_{w,r}$, whose projection on the horizontal axis contains x. The ratio of the slope of d_{n+1} to the slope of d_n is either w/r or $(1 - w)/(1 - r)$. Since both numbers differ from 1, the slope of d_n cannot converge to a finite, positive limit. Hence, $S_{w,r}$ does not have a finite, positive derivative at x.

Suppose μ assigns measure 1 neither to a point, nor to the main diagonal, and $G \in \Delta$ is not the uniform distribution. Are P_μ-almost all F singular with respect to G? In case μ assigns measure $\frac{1}{2}$ to each of the points $(\frac{1}{3}, \frac{1}{2})$ and $(\frac{2}{3}, \frac{1}{2})$?

The strict singularity of $S_{w,r}$ has various other generalizations (sections 6, 7, 8). Let K be a subset of the unit square. Say $F \in \Delta$ is *K-constructible* if its graph can be obtained by the CONSTRUCTION, with this constraint: at each stage, each point selected from each rectangle is in the positive, affine image of K in that rectangle. Of course, no base probability is involved in this definition. If, for every strictly convex $F \in \Delta$, and strictly concave $G \in \Delta$, there is a point $(x, y) \in K$ with $F(x) < y < G(x)$, then K is *tangent* to the main diagonal. The main result of section 6 is the following. If K is tangent to the main diagonal, then there is a K-constructible distribution function equivalent to Lebesgue measure; otherwise, every K-constructible distribution function is purely singular. We do not know necessary and sufficient conditions on K for each K-constructible distribution function to be strictly singular, but as section 7 shows, this condition is sufficient: for some strictly convex $F \in \Delta$, every point of K is on or below the graph of F. If K is a compact subset of the interior of S, and is disjoint from the main diagonal, no K-constructible distribution function has even a finite, positive, one-sided derivative anywhere (7.2). We do not know necessary and sufficient conditions on K for this to hold.

Let F and G be distribution functions: F is *strictly singular* with respect to G provided there is no x for which the ratio of $F(x + h) - F(x)$ to $G(x + h) - G(x)$ converges to a finite, positive limit as h tends to 0. Section 8 proves the following. Let $0 < r < 1$, and let μ and ν be distinct base probabilities, each assigning measure 1 to the vertical line segment $x = r$, $0 < y < 1$. Then there are Borel subsets C and D of Δ, with $P_\mu(C) = P_\nu(D) = 1$, such that F is strictly singular with respect to G for all $F \in C$ and $G \in D$. In particular, $P_\mu \neq P_\nu$. It is likely that, unless μ concentrates on the main diagonal, P_μ determines μ.

A probability P on Δ determines an *average distribution function* F_P according to the relation

$$F_P(z) = \int_{G \in \Delta} G(z) P(dG). \tag{1.1}$$

The average F_{P_μ}, or F_μ for short, is a fixed point for T_μ (section 9), where T_μ is this mapping of Δ into itself. For $G \in \Delta$, $T_\mu G$ is the distribution function of a point v chosen from I according to this mechanism: choose a point u at random from I according to G, and independently a point (x, y) at random from S according to μ; then v is xu with probability y, and $u + x(1 - u)$ with probability $1 - y$. Consequently, as [7] implies: T_μ has F_μ for its unique fixed point, and for $G \in \Delta$, $(T_\mu)^n G$ converges to F_μ uniformly as $n \to \infty$; F_μ is continuous, strictly monotone, and either purely singular or absolutely continuous.

Here is a more special result. Let $\theta(\mu, x)$ be the conditional μ-expectation of the second coordinate given that the first coordinate is x, and let μ_h be the projection of μ on the horizontal axis. Then F_μ is Lebesgue measure if and only if $\theta(\mu, x) = x$ for μ_h-almost all x (9.28). We conjecture that unless F_μ is the uniform distribution, it determines $\theta(\mu, \cdot)$ and μ_h.

In certain special cases, it is possible to compute F_μ explicitly. If μ concentrates on the vertical line segment $x = r$, $0 < y < 1$, and has mean (r, w), then $F_\mu = S_{w,r}$. If μ is the uniform distribution on S, or on the horizontal line segment $0 < x < 1$, $y = \frac{1}{2}$, then $F_\mu(z) = 2\pi^{-1} \arcsin z^{1/2}$. This contrasts with the fact that F_μ is the uniform distribution if μ is uniform on the vertical line segment $x = \frac{1}{2}$, $0 < y < 1$.

We do not know which $F \in \Delta$ are of the form F_μ for some base probability μ, nor whether distinct F_μ can agree on an interval, nor when F_μ's are equivalent or singular.

2. The definition of the prior P_μ

Formalizing the CONSTRUCTION of section 1 seems to require a fair amount of notation: I is the closed unit interval $[0, 1]$, and S is the closed unit square $I \times I$. For any sets X and Y, X^Y is the set of all functions from Y to X. For $n = 0, 1, \cdots$, B_n is the set of all n-tuples of 0's and 1's; the only element of B_0 is the (empty) 0-tuple $\varnothing$. For each $b \in B_n$, b followed by 0, namely $b0$, is in B_{n+1}, as is $b1$. Let $B = \bigcup_{n=0}^\infty B_n$.

For any closed subinterval J of I, $\langle J \rangle$ is the linear mapping of I onto J which carries 0 to the left endpoint of J. For $t \in I^B$ and $b \in B$, define the real number $q(b, t)$ and the closed interval $J(b, t)$ inductively: $J(\varnothing, t)$ is I; $q(b, t)$ is the image of $t(b)$ under $\langle J(b, t) \rangle$; $J(b0, t)$ is the image of the closed interval $[0, t(b)]$ under $\langle J(b, t) \rangle$; and $J(b1, t)$ is the image of $[t(b), 1]$ under $\langle J(b, t) \rangle$.

Of course, $\tau \in S^B$ can be identified with a pair $(\tau_1, \tau_2) \in I^B \times I^B$ by the relation $\tau(b) = (\tau_1(b), \tau_2(b))$ for all $b \in B$. For $\tau \in S^B$ and $b \in B$, define $p(b, \tau) \in S$ and $r(b, \tau)$, a closed subrectangle of S, by

$$p(b, \tau) = (q(b, \tau_1), q(b, \tau_2)), \tag{2.1}$$

$$r(b, \tau) = J(b, \tau_1) \times J(b, \tau_2). \tag{2.2}$$

Let

$$M_n(\tau) = \bigcup \{r(b, \tau) : b \in B_n\} \tag{2.3}$$

and

$$(2.4)\qquad M_\infty(\tau) = \bigcap_{n=0}^{\infty} M_n(\tau).$$

Let Δ be the set of distribution functions on I, normalized to be 1 at 1 and continuous from the right on the half-open interval $[0, 1)$. If $F \in \Delta$, its *solid graph* ηF is the smallest closed subset of S whose intersection with every vertical line is convex, which contains $(0, 0)$, and which includes the customary graph of F. Let T be the set of all $\tau \in S^B$ such that $M_\infty(\tau)$ is the solid graph of some $F \in \Delta$, and let $M(\tau)$ be this F.

For any sets X and Y, if X is endowed with a topology (σ-field), then X^Y is given the product topology (σ-field). If μ is a probability on (a σ-field of subsets of) X, the *power probability* μ^Y on (the product σ-field of) X^Y is defined by this property: as y ranges over Y, the coordinate functions $\hat{y}$, with $\hat{y}(\omega) = \omega(y) \in X$ for all $\omega \in X^Y$, are independent under μ^Y and have common distribution μ. Endow Δ with the weak* topology, that is, the smallest topology such that $F \to \int_I f(x)F(dx)$ is continuous for each continuous, real-valued function f on I. Endow Δ with the σ-field generated by this compact, metrizable topology.

(2.5) LEMMA. *T is a Borel subset of S^B, and M is a Borel measurable map of T into Δ.*

PROOF. Plainly, η is continuous and $1 - 1$ from Δ into 2^S, the set of nonempty, closed subsets of S; for a discussion of 2^S, see ([12], section 28) or ([19], section 15 and [20], section 38). Consequently, η^{-1} is continuous on its domain $\eta\Delta$, and therefore Borel measurable. For $n < \infty$, each M_n is continuous; so M_∞ is upper semicontinuous, and therefore Borel measurable [18]. Finally, $T = M_\infty^{-1}\eta\Delta$ and $M = \eta^{-1}M_\infty$. ◆

(2.6) DEFINITION. If the probability μ on S assigns measure 0 to the corners $(0, 0)$ and $(1, 1)$ of S, then μ is a *base probability*.

(2.7) LEMMA. *If μ is a base probability, then $\mu^B(T) = 1$.*

PROOF. Let $A_n(\tau)$ be the planar Lebesgue measure of $M_n(\tau)$, $n = 0, 1, \cdots, \infty$; then $\tau \in T$ if and only if $A_\infty(\tau) = 0$, or equivalently, the monotone decreasing sequence $A_0(\tau), A_1(\tau), \cdots$ converges to 0. But $A_n(\tau)$ does converge to 0 for μ^B-almost all τ because its μ^B-expected value is ρ^n with $0 < \rho < 1$. In more detail, let

$$(2.8)\qquad \rho = \int_S [xy + (1 - x)(1 - y)]\mu(dx, dy).$$

Then ρ is the μ^B-expectation of A_1; and if $n \geq 1$, the conditional μ^B-expectation of A_{n+1}, given $\tau(b)$ for $b \in \bigcup_{j=0}^{n-1} B_j$, is $\rho A_n(\tau)$. ◆

Each probability Q on S^B is transformed by M into the subprobability QM^{-1} on Δ. Plainly, QM^{-1} is a probability if and only if $Q(T) = 1$. The principal concept of this paper can now be introduced.

(2.9) DEFINITION. For each base probability μ on S, the *geometric prior* P_μ is the probability $\mu^B M^{-1}$ on Δ.

(2.10) NOTATION. Throughout the rest of this paper, except in (9.2) to (9.11), μ is a base probability.

We do not know an abstract characterization of the set of priors P_μ as μ ranges over the base probabilities.

3. The support of P_μ

This section gives a necessary and sufficient condition for the support of P_μ to be all of Δ (3.6). Some readers may prefer to skip this section, which is not relied upon in future sections. As usual, the *support* of a probability on a compact metric space is the smallest closed subset of probability 1.

(3.1) DEFINITION. A distribution function F is *K-constructible* if there is a $\tau \in T$ with $\tau(b) \in K$ for all $b \in B$ and $M(\tau) = F$. If K is the support of a base probability μ, then *μ-constructible* will mean K-constructible.

(3.2) LEMMA. *The support of P_μ is the closure of the μ-constructible distribution functions.*

Because M is continuous on T, lemma (3.2) is an immediate consequence of (3.3) and (3.4), whose easy proofs are omitted.

(3.3) LEMMA. *Let φ be a measurable map of a probability triple $(\Omega, \mathfrak{F}, P)$ into a compact metric space X. If $W \in \mathfrak{F}$ has P-measure 1, the closure of $\varphi(W)$ includes the support of $P\varphi^{-1}$.*

The next lemma uses the notation φ_D for the restriction of a function φ to a part D of its domain.

(3.4) LEMMA. *Let Ω and X be compact metric spaces, φ measurably defined from part of Ω to X, P a probability on Ω, K the support of P, and D a Borel subset of Ω having P-measure 1 such that φ_D is continuous. Then $\varphi(D \cap K)$ is included in the support of $P\varphi^{-1}$.*

Of course, (3.2) implies the following.

(3.5) COROLLARY. *If the support of a base probability μ includes that of another ν, then the support of P_μ includes that of P_ν.*

Plainly, if μ and ν have the same support, so do P_μ and P_ν. If K is any nonempty compact subset of S which does not contain $(0, 0)$ or $(1, 1)$ as isolated points, then all base probabilities whose support is K lead to priors with the same support, $\sum(K)$. This section gives a necessary and sufficient condition for $\sum(K)$ to be all of Δ.

To state the condition, call K *horizontal* at $(0, 0)$ if for each $\epsilon > 0$ there is an $(x, y) \in K$ with $0 < x < \epsilon$ and $x^{-1}y < \epsilon$. Likewise, K is *vertical* at $(0, 0)$ if for each $\epsilon > 0$ there is an $(x, y) \in K$ with $0 < y < \epsilon$ and $x^{-1}y > \epsilon^{-1}$. Call $F \in \Delta$ *elementary* if its solid graph ηF consists of a finite number of line segments, each of which is horizontal or vertical. If K and ηF are both horizontal or both vertical at $(0, 0)$, then ηF is *tangent* to K at $(0, 0)$. The analogous definition for ηF to be *tangent* to K at $(1, 1)$ is omitted.

It is now easy to state a necessary and sufficient condition for $\sum(K)$ to be Δ.

(3.6) THEOREM. *Let K be a nonempty compact subset of S which does not contain*

$(0, 0)$ or $(1, 1)$ as isolated points. Then $\sum(K) = \Delta$ if and only if the solid graph of every elementary distribution function tangent to K neither at $(0, 0)$ nor at $(1, 1)$ intersects K at a point other than $(0, 0)$ and $(1, 1)$.

(3.7) COROLLARY. *If K is tangent to all elementary distribution functions, then $\sum(K) = \Delta$.*

(3.8) COROLLARY. *If K contains neither $(0, 0)$ nor $(1, 1)$, then $\sum(K) = \Delta$ if and only if K intersects the solid graph of every distribution function.*

Write $s(1)$ for the first coordinate of $s \in S$, and $s(2)$ for the second. Here are five examples of K with $\sum(K) = \Delta$ (figures 3.1–3.5).

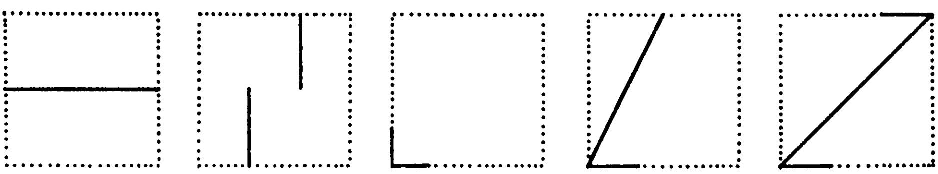
(3.9) EXAMPLE. The horizontal line segment $\{s\colon s \in S \text{ and } s(2) = \frac{1}{2}\}$.

(3.10) EXAMPLE. This union of two line segments: $\{s\colon s \in S \text{ and } s(1) = \frac{1}{3}, 0 \le s(2) \le \frac{1}{2}\} \cup \{s\colon s \in S \text{ and } s(1) = \frac{2}{3}, \frac{1}{2} \le s(2) \le 1\}$.

(3.11) EXAMPLE. This union of two line segments: $\{s\colon s \in S \text{ and } s(1) = 0, 0 \le s(2) \le \frac{1}{4}\} \cup \{s\colon s \in S \text{ and } 0 \le s(1) \le \frac{1}{4}, s(2) = 0\}$.

(3.12) EXAMPLE. This union of two line segments: $\{s\colon s \in S \text{ and } s(2) = 2s(1) \text{ for } 0 \le s(1) \le \frac{1}{2}\} \cup \{s\colon s \in S \text{ and } 0 \le s(1) \le \frac{1}{3}, s(2) = 0\}$.

(3.13) EXAMPLE. This union of three line segments: $\{s\colon s \in S \text{ and } s(2) = s(1)\} \cup \{s\colon s \in S \text{ and } 0 \le s(1) \le \frac{1}{3}, s(2) = 0\} \cup \{s\colon s \in S \text{ and } \frac{2}{3} \le s(1) \le 1, s(2) = 1\}$.



3.1 3.2 3.3 3.4 3.5

FIGURES

The rest of this section is devoted to the proof of the theorem.

PROOF OF (3.6). The "only if" is settled by proving a little more.

(3.14) *An elementary distribution function F tangent to K at neither $(0, 0)$ nor $(1, 1)$, and intersecting K at no points other than $(0, 0)$ and $(1, 1)$, is not in $\sum(K)$.*

Suppose F horizontal at $(0, 0)$ and $(1, 1)$, for no new difficulty arises in the other cases.

For $G \in \Delta$, let G^a be the set of points above the graph of G, namely, the set of $s \in (S - \eta G)$ with $s(2) > G(s(1))$. Similarly, G^b is the set of points below the graph of G, namely, the set of $s \in (S - \eta G)$ with $s(2) < G(s(1))$.

By elementary continuity considerations there are positive numbers β, a_1, a_2, b_1, b_2, c_1, c_2 less than 1 and continuous distribution functions F_1, F_2, G_1, G_2 satisfying (3.15) to (3.19), as depicted in figure 3.6.

(3.15) $K \subset F_1^a \cup F_2^b \cup \{(0, 0)\} \cup \{(1, 1)\}$.

(3.16) $b_1 < c_2 < a_2$ and $a_1 < c_1 < b_2$.

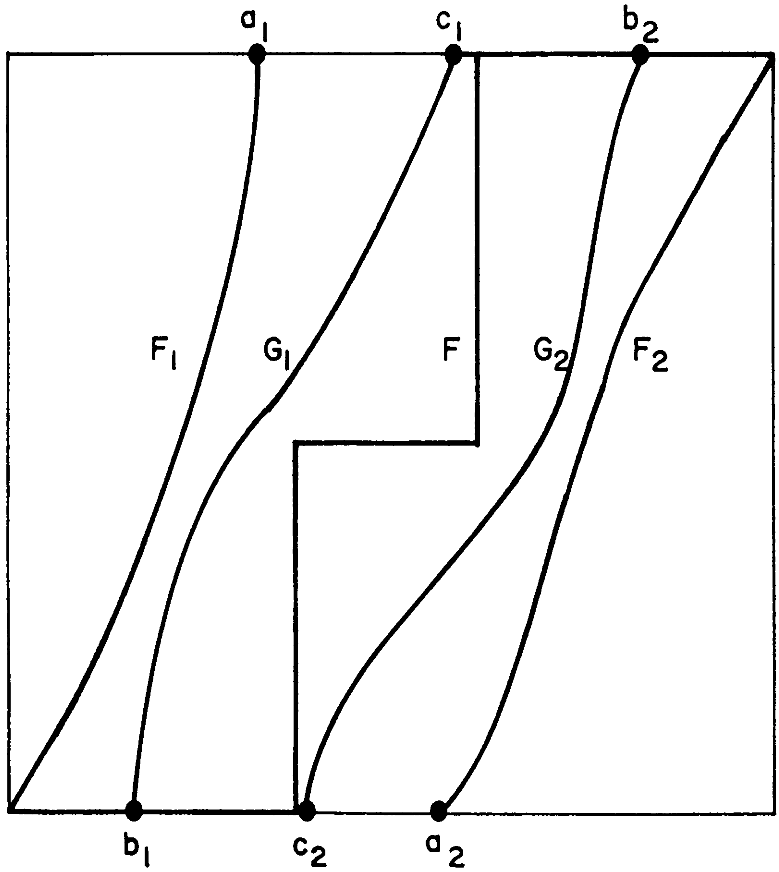


FIGURE 3.6

(3.17) F_1 is linear with finite, positive slope α on $[0, \beta]$, and F_2 is linear with slope α on $[1 - \beta, 1]$.

(3.18) $F_1(a_1) = G_1(c_1) = F(b_2) = G_2(b_2) = 1$, and
$G_1(b_1) = F(b_1) = G_2(c_2) = F_2(a_2) = 0$.

(3.19) On $(0, a_1]$, $F_1 > G_1$ and is strictly monotone;
on $(b_1, c_1]$, $G_1 > F$ and is strictly monotone;
on $[c_2, b_2)$, $G_2 < F$ and is strictly monotone;
on $[a_2, 1)$, $F_2 < G_2$ and is strictly monotone.

For each $\epsilon > 0$, let $G_{1\epsilon} = \max \{\epsilon, G_1\}$; and let $G_{2\epsilon} = \min \{1 - \epsilon, G_2\}$ on $[0, 1)$, $G_{2\epsilon}(1) = 1$. Let $V_\epsilon = (G_{1\epsilon}^b \cap G_{2\epsilon}^a) \cup \{s: s \in S \text{ and } s(1) = 0, 0 \leq s(2) < \epsilon\} \cup \{s: s \in S \text{ and } 0 \leq s(1) < c_2, s(2) = 0\} \cup \{s: s \in S \text{ and } s(1) = 1, 1 - \epsilon < s(2) \leq 1\} \cup \{s: s \in S \text{ and } c_1 < s(1) \leq 1, s(2) = 1\}$ (see figure 3.7). Let $V_\epsilon^* = \{G: G \in \Delta \text{ and } \eta G \in V_\epsilon\}$. Plainly, V_ϵ^* is an open neighborhood of F. The next step is to prove:

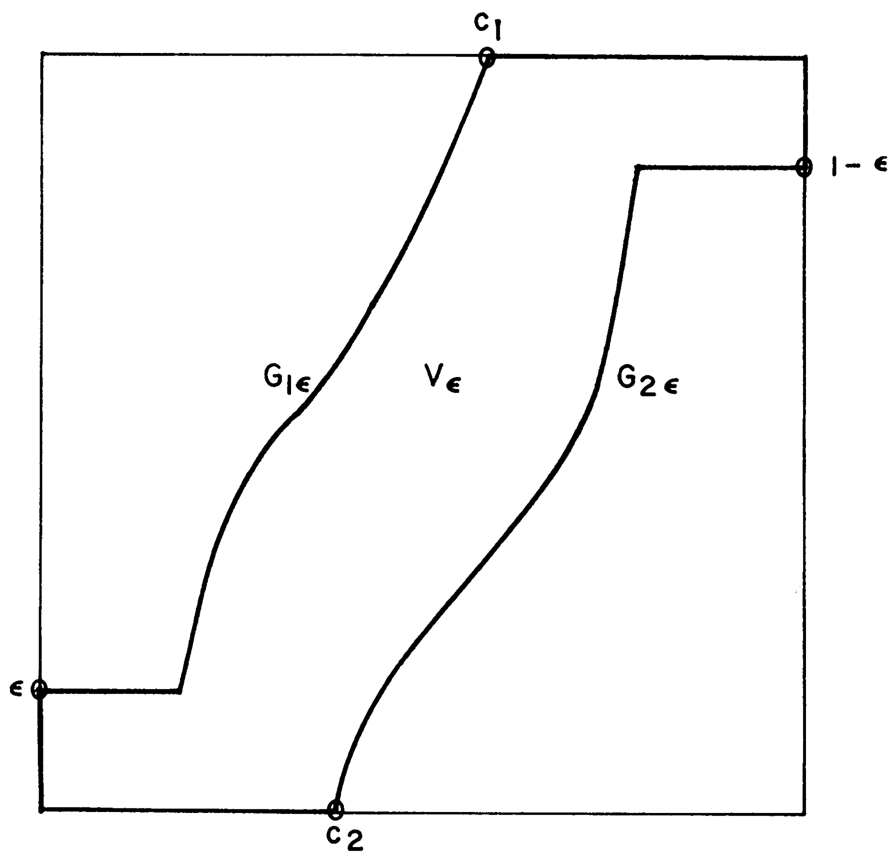


FIGURE 3.7

(3.20) *For small enough* ϵ, *there are no K-constructible distribution functions in* V_ϵ^*. Lemma (3.2) will then apply and give (3.14).

If s and t are in S with $s(i) \leq t(i)$, $i = 1, 2$, then $s \times t$ is the closed subrectangle of S whose lower left corner is s and upper right corner is t. The *positive, affine map* $A_{s\times t}$ of S onto $s \times t$ is the map $(u, v) \rightarrow (au + b, cv + d)$ with a and c nonnegative which sends S onto $s \times t$.

There is a positive $\rho < \frac{1}{2}$ such that for s and t in S with $s(i) < \rho$ and $1 - t(i) < \rho$ for $i = 1, 2$:

(3.21) $$A_{s\times t}F_1^a \subset G_1^a;$$

(3.22) $$A_{s\times t}F_2^b \subset G_2^b;$$

and

(3.23) the image under $A_{s\times t}$ of every line segment has at least $\frac{1}{2}$ its original length.

Recall the meanings of α and β from (3.17) and choose $\epsilon > 0$ so small that

(3.24) $$\epsilon < \rho,$$

(3.25) $$\delta < \rho,$$

and

(3.26) $$\epsilon^2 + \delta^2 < \tfrac{1}{4}\beta^2(1 + \alpha^2),$$

where $\delta = (\epsilon/\alpha(1 - 2\epsilon))$. Let T_0 be the closed triangle with vertices $(0, 0)$, $(0, \epsilon)$, (δ, ϵ); and T_1 the closed triangle with vertices $(1, 1)$, $(1, 1 - \epsilon)$, $(1 - \delta, 1 - \epsilon)$.

Suppose $\tau \in S^B$ has $\tau(b) \in K$ for all $b \in B$ and $M_\infty(\tau) \subset V_\epsilon$. Once it is argued that

(3.27) $$\tau \notin T,$$

relation (3.20) will be established. More than (3.27) will be proved:

(3.28) For all $n \geq 0$ there is a (necessarily unique) $b_n \in B_n$ with the lower left corner of $r(b_n, \tau)$ in T_0 and the upper right corner in T_1.

Plainly, (3.28) holds for $n = 0$ and $n = 1$. Suppose it holds for $n = k$. To prove it for $n = k + 1$ requires only the verification that $p(b_k, \tau) \in T_0 \cup T_1$. If $\tau(b_k) \in F_1^a$, it will be seen that $p(b_k, \tau) \in T_0$. The case $\tau(b_k) \in F_2^b$ is omitted. Abbreviate R for $r(b_k, \tau)$ and σ_1 for the segment of ηF_1 over $[0, \beta]$ (see (3.17)).

Since $p(b_k, \tau)$ is in $A_R F_1^a$, it is above ηG_1 (see figure 3.8 and use the induction

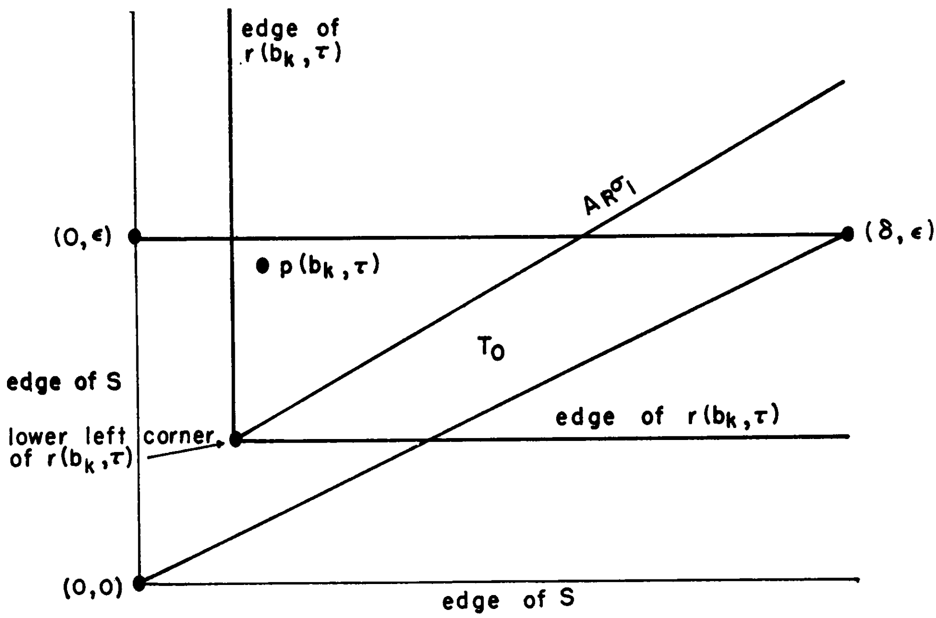


FIGURE 3.8

hypothesis, (3.24), (3.25)). Since it is in V_ϵ, it must be below the line $s(2) = \epsilon$; (figures 3.6 and 3.7). But the length of $A_R\sigma_1$ exceeds the diameter of T_0 (the induction hypothesis and (3.24) to (3.26)). Thus $p(b_k, \tau)$ is in the closed triangle bounded by $s(2) = \epsilon$, the left edge of $r(b_k, \tau)$, and $A_R\sigma_1$. To finish the argument, it is enough to prove that the slope of $A_R\sigma_1$ is not less than the slope of the hypotenuse of T_0, which is $\alpha(1 - 2\epsilon)$. Among all rectangles $s_0 \times s_1$ with $s_0 \in T_0$ and $s_1 \in T_1$, the slope of $A_{s_0 \times s_1}\sigma_1$ is minimized with $s_0 = (\epsilon, 0)$ and $s_1 = (1, 1 - \epsilon)$, and is then precisely $\alpha(1 - 2\epsilon)$. This completes the induction, and with it the proof of the "only if" part of the theorem.

Turn now to the "if" part of (3.6). The *order* of an elementary $F \in \Delta$ is the minimal number of line segments of which ηF is composed. It will be proved by induction on the order that every elementary $F \in \Delta$ is in $\sum(K)$. The two elementary F's of order 2 are in $\sum(K)$ by a relatively easy argument which is omitted. Suppose that for some $k \geq 3$, every elementary F of order less than k is in $\sum(K)$. To see that every F of order k is in $\sum(K)$, consider two cases.

Case 1: *F is not tangent to K.* Suppose first that k is odd, so the initial and final segments of ηF are parallel, say vertical. Let $K_0 = \{s\colon s \in K, s(1) = 0, s(2) > 0\}$ and $K_1 = \{s\colon s \in K, s(1) = 1, s(2) < 1\}$. Suppose that K_0 and K_1 are nonempty, for no new difficulty arises in the other case. Let p be the lowest point of K_0 and q the highest point of K_1. These points exist because F is not tangent to K. Suppose that F were not in $\sum(K)$. Then ηF cannot intersect K at an interior point of S. For let $s \in \eta F$ be an interior point of S. The part of ηF between $(0, 0)$ and s, when rescaled so that s becomes $(1, 1)$, is the solid graph of an elementary distribution function, namely $\eta^{-1}A^{-1}_{(0,0)\times s}\eta F$. This distribution function has order less than k, so is in $\sum(K)$ by the induction hypothesis. Similarly, $\eta^{-1}A_{s\times(1,1)}\eta F$ is in $\sum(K)$. If s were in K, F would plainly be in $\sum(K)$. Either $p \in \eta F$ or $q \in \eta F$ (by the condition of the theorem); say $p \in \eta F$.

It is now convenient to define an element τ of K^B, and a sequence $R_0, R_1, \cdots$ of subrectangles of S; this requires the simultaneous definition of a sequence $b_0, b_1, \cdots$ of elements of S. Let $b_0 = \varnothing$. For $n \geq 0$, let $b_{n+1} = b_n1$ and $\tau(b) = p$ for $b \in B_n$, as long as $p(b_j, \tau)$ is in the initial segment of ηF for every $j \leq n$. Plainly, this part of the defining procedure cannot continue indefinitely. Let n_0 be the last n such that $\tau(b)$ has been defined for $b \in B_n$, and let $R_0 = p(b_{n_0}, \tau) \times (1, 1)$. Since $F \notin \sum(K)$, $(\eta^{-1}A_{R_0}^{-1}\eta F)(0) < p(2)$ and $A_{R_0}^{-1}\eta F$ does not intersect K at an interior point of S. Because ηF is not tangent to K, neither is $A_{R_0}^{-1}\eta F$. By the condition of the theorem, $q \in A_{R_0}^{-1}\eta F$. The definition of τ can now be continued for $n > n_0$, by setting $b_{n+1} = b_n0$ and $\tau(b) = q$ for $b \in B_n$, as long as $p(b_j, \tau)$ is in the final segment of ηF for $n_0 + 1 \leq j \leq n$. This part of the defining procedure also cannot continue indefinitely. Let n_1 be the last n such that τ has now been defined for $b \in B_n$, and let $R_1 = p(b_{n_0}, \tau) \times p(b_{n_1}, \tau)$. Now return to selecting p, and so on, thereby defining R_2, and so on, completing the definition of τ and the sequence $R_0, R_1, \cdots$ of rectangles.

Plainly, the height of R_n converges to 0. For $k = 3$ there is a contradiction, because $\tau \in K^B \cap T$ and $M(\tau) = F$. Also for $k > 3$ there is a contradiction, because the height of R_n is bounded below by the F-measure of the open interval $(0, 1)$.

For even k, with the initial segment of ηF vertical and the final segment horizontal, replace q by the rightmost point of $\{s\colon s \in K,\ s(1) < 1,\ s(2) = 1\}$, and use a similar argument.

Case 2: *F is tangent to K*; say at $(0, 0)$ where both are horizontal, for definiteness. Let x be the right endpoint of the initial segment of ηF; so $F(y) > 0$ for $y \geq x$. If $\epsilon > 0$, there is a $G \in \sum(K)$ whose graph is within ϵ of ηF. To see this, choose an $(r, w) \in K$ such that $0 < r < \epsilon$, and if n is the least natural number with $r' > x$, then $w' < \min\{\epsilon, F(x)\}$, where $r' = 1 - (1 - r)^n$ and $w' = 1 - (1 - w)^n$. If $F_1 = \eta F \cup \{s\colon s \in S \text{ and } s(1) = r',\ w' \leq s(2) \leq F(r')\}$, then $\eta^{-1}A^{-1}_{(r',w')\times(1,1)}F_1$ is an elementary distribution function of order less than k, so is in $\sum(K)$. The balance of the argument is routine. ♦

4. Continuity and monotoneity

Call $F \in \Delta$ *continuous* if $F(0) = 0$ and F is continuous on I in the usual sense. The main result of this section is the following theorem.

(4.1) THEOREM. *P_μ assigns probability 1 to the set of continuous and strictly increasing distribution functions if and only if μ assigns probability 1 to the interior of the unit square.*

The first lemma uses notation introduced in section 2.

(4.2) LEMMA. *If a probability θ on I assigns positive probability to the interior of I, then*

$$\lim_{n\to\infty} \max_{b\in B_n} \text{length of } J(b, t) = 0 \tag{4.3}$$

for θ^B-almost all $t \in I^B$.

PROOF. Each $r \in I$ divides I into two intervals, of length r and $1 - r$ respectively. The θ-expectation of the sum of the squares of these lengths is $\rho = \int_I [r^2 + (1 - r)^2]\theta(dr) < 1$. Let $L_n(t)$ be the sum of the squares of the lengths of the 2^n intervals $J(b, t)$ for $b \in B_n$. The θ^B-expectation of L_n is ρ^n, which converges to 0. Since $L_{n+1} \leq L_n$, $L_n(t)$ converges to 0 for θ^B-almost all t. For each such t, equation (4.3) holds. ♦

(4.4) LEMMA. *A necessary and sufficient condition for P_μ-almost all distribution functions to be continuous is:* (i) *μ assigns probability 0 to the vertical edges of S, and* (ii) *μ assigns positive probability to the interior of S.*

PROOF. The condition is plainly necessary. To prove sufficiency, apply (4.2) with the projection of μ on the vertical axis for θ, and conclude from (ii) that for μ^B-almost all τ,

(4.5) the maximum of the heights of the 2^n rectangles comprising $M_n(\tau)$ converges to 0.

In view of (i), for μ^B-almost all τ,

(4.6) for each n, each vertical line intersects at most two of the rectangles comprising $M_n(\tau)$.

For τ satisfying (4.5) and (4.6), the intersection of each vertical line with $M_n(\tau)$

is an interval whose length shrinks to 0 as $n \to \infty$, that is, each vertical line intersects $M_\infty(\tau)$ in at most a single point. ◆

Lemma (4.8) can be given a proof similar to that of (4.4), but it also follows from (4.4) by a formal use of duality.

Let σ_0 be the identity mapping of S onto S; while σ_1 is the symmetry of S which sends (x, y) to (y, x), and σ_2 sends (x, y) to $(1 - x, 1 - y)$. The four symmetries σ_0, σ_1, σ_2, and $\sigma_1\sigma_2$ constitute the *Δ-group*. Each σ in the Δ-group carries the solid graph of an $F \in \Delta$ into the solid graph of a new element σ^*F of Δ. Plainly, σ^* is a homeomorphism of Δ.

If $(\Omega, \mathfrak{F}, P)$ is a probability triple, and f is a $1 - 1$ bimeasurable mapping of Ω onto itself, the probability Pf is defined by the relation $(Pf)(A) = P(fA)$. With the help of this notation, it is easy to state the duality principle.

(4.7) LEMMA. *If σ is in the Δ-group, then $P_\mu\sigma^* = P_{\mu\sigma}$.*

PROOF. The proof is easy. ◆

(4.8) LEMMA. *A necessary and sufficient condition for P_μ-almost all distribution functions to be strictly increasing is that μ assigns probability 0 to the horizontal edges of S, and positive probability to the interior of S.*

PROOF. An $F \in \Delta$ is strictly increasing if and only if σ_1^*F is continuous. Apply (4.4) and (4.7). ◆

Plainly, (4.1) is an immediate consequence of (4.4) and (4.8).

(4.9) THEOREM. *A necessary and sufficient condition for P_μ-almost all distribution functions to be purely discrete is: either the horizontal edges of S have μ-probability 1 or a vertical edge has positive μ-probability.*

PROOF. Necessity is obvious by (4.4). Turn to sufficiency. If the horizontal edges of S have μ-probability 1, P_μ-almost all F obviously have a jump of size 1. Suppose a vertical edge has positive μ-probability. For each $F \in \Delta$, let $d(F)$ be the sum of the jumps of F; let D be the P_μ-expectation of d, or equivalently the μ^B-expectation of $d(M)$; and let $E(s)$ be the conditional μ^B-expectation of $d(M)$ given $\tau(\varnothing) = s$. Plainly,

$$D = \int_S E(s)\mu(ds) \leq 1; \tag{4.10}$$

$$\begin{aligned} E(s) &= s(2) + (1 - s(2))D && \text{for} \quad s(1) = 0 \\ &= 1 - s(2) + s(2)D && \text{for} \quad s(1) = 1 \\ &= D && \text{for} \quad 0 < s(1) < 1, \end{aligned} \tag{4.11}$$

where $s = (s(1), s(2))$.

Therefore $D \leq E(s)$ for all $s \in S$. Hence (4.10) implies $D = E(s)$ for μ-almost all s, and in particular for at least one s other than $(0, 0)$ or $(1, 1)$ on a vertical edge of S. Therefore $D = 1$, so $d(F) = 1$ for P_μ-almost all F. ◆

A distribution function is *purely flat* if its support has Lebesgue measure 0.

(4.12) THEOREM. *A necessary and sufficient condition for P_μ-almost all distribution functions to be purely flat is that either the vertical edges of S have μ-probability 1 or a horizontal edge has positive μ-probability.*

PROOF. Apply (4.7) and (4.9). ◆

The following lemma is probably known.

(4.13) LEMMA. *Let* (u_n, ϵ_n): $n \geq 1$ *be a sequence of pairs of positive real numbers with* $u_n \to 0$, $\epsilon_n \to 0$, *and* $\sum u_n = \infty$. *There is a subset* N *of natural numbers with* $\sum_{n \in N} u_n = \infty$ *and* $\sum_{n \in N} \epsilon_n u_n < \infty$.

PROOF. Find a sequence of positive integers $j_1, k_1, j_2, k_2, \cdots$ with

$$j_{i+1} > j_i + k_i, \tag{4.14}$$

$$\frac{1}{i} \leq U_i \leq \frac{2}{i}, \tag{4.15}$$

and

$$\sup_{n \geq j_i} \epsilon_n \leq \frac{1}{i}, \tag{4.16}$$

where

$$U_i = \sum_{n=j_i}^{j_i+k_i} u_n. \tag{4.17}$$

Let N_i be the set of $k_i + 1$ numbers $\{j_i, j_i + 1, \cdots, j_i + k_i\}$, and let $N = \bigcup_{i=1}^{\infty} N_i$. Then

$$\sum_{n \in N} u_n = \sum_{i=1}^{\infty} U_i \geq \sum_{i=1}^{\infty} \frac{1}{i} = \infty, \tag{4.18}$$

and

$$\begin{aligned} \sum_{n \in N} \epsilon_n u_n &= \sum_{i=1}^{\infty} \sum_{n=j_i}^{j_i+k_i} \epsilon_n u_n \\ &\leq \sum_{i=1}^{\infty} (\sup_{n \geq j_i} \epsilon_n) U_i \\ &\leq \sum_{i=1}^{\infty} \frac{1}{i} \frac{2}{i} < \infty. \end{aligned} \tag{4.19}$$

◆

(4.20) THEOREM. *Let* K *be a compact subset of* S *containing no point of the boundary of* S, *except possibly* $(0, 0)$ *or* $(1, 1)$. *If* K *is tangent to a vertical edge of* S, *some* K*-constructible distribution function is purely flat; otherwise each* K*-constructible distribution function is strictly increasing.*

For the definition of "K-constructible," see (3.1).

PROOF. Suppose that K is tangent to the left edge of S. That is, there are points $(x_n, y_n) \in K$ with $x_n \to 0$ and $x_n^{-1} y_n \to \infty$. By (4.13), suppose without loss of generality that $\sum x_n < \infty$ and $\sum y_n = \infty$. Define $\tau \in S^B$ by the relation $\tau(b) = (x_n, y_n)$ for all $b \in B_n$.

Since $\prod (1 - y_n)$ is 0 and $y_n \to 0$, $\tau \in T$. Moreover, the distribution function $M(\tau)$ is flat at least on the interval of length $\prod_{n=0}^{\infty} (1 - x_n)$ whose right endpoint is 1.

Apply this reasoning to the rectangles comprising $M_n(\tau)$. If one of these rectangles R has width w, then $M_\infty(\tau)$ includes a horizontal line segment of length $w \prod_{j=n}^{\infty} (1 - x_j)$, at the extreme right of the upper edge of R. The sum of the widths of the rectangles comprising $M_n(\tau)$ is 1; so $M_\infty(\tau)$ includes a set of horizontal line segments of total length $\prod_{j=n}^{\infty} (1 - x_j)$. This expression converges to 1 as $n \to \infty$, which proves the first assertion.

Suppose K is not tangent to a vertical edge of S. Then there is a positive, finite c such that $y \leq cx$ for all $(x, y) \in K$. If $(x_i, y_i) \in K$ and $\sum y_i = \infty$, then $\sum x_i = \infty$. Consequently,

(4.21) $\quad \prod (1 - y_i) = 0$ and all $y_i < 1$ imply $\prod (1 - x_i) = 0$.

Likewise,

(4.22) $\quad \prod y_i = 0$ and all $y_i > 0$ imply $\prod x_i = 0$.

Suppose for all $b \in B$ that $\tau(b) \in K$ and $\tau(b)$ is neither $(0, 0)$ nor $(1, 1)$. Let $\{b_n: n \geq 0\}$ be a *path* through B; that is, b_0 is $\varnothing$ and b_{n+1} is b_n0 or b_n1. If the height of $r(b_n, \tau)$ converges to 0, so does its width, by (4.21) and (4.22). That is, $M_\infty(\tau)$ includes no horizontal line segments. Moreover,

(4.23) If K is a subset of S, and F is a K-constructible distribution function, there is a $\tau \in S^B$ with $F = M(\tau)$ and, for all $b \in B$, $\tau(b)$ is in K but is neither $(0, 0)$ nor $(1, 1)$. ♦

(4.24) THEOREM. *Let K be a compact subset of S containing no point of the boundary of S, except possibly $(0, 0)$ or $(1, 1)$. If K is tangent to a horizontal edge of S, some K-constructible distribution function is purely discrete; otherwise, each K-constructible distribution function is continuous.*

PROOF. Apply (4.7) and (4.20). ♦

5. Almost all distribution functions are strictly singular

A distribution function is *strictly singular* if it has a finite, positive derivative nowhere. This section is devoted to proving the following theorem.

(5.1) THEOREM. *Unless μ assigns measure 1 to the main diagonal of the unit square, P_μ-almost all distribution functions are strictly singular.*

Of course, if $F \in \Delta$ is strictly singular, it is singular with respect to Lebesgue measure. Thus, if μ does not assign measure 1 to the main diagonal, P_μ-almost all $F \in \Delta$ are singular with respect to Lebesgue measure. If $G \in \Delta$, when are P_μ-almost all F singular with respect to G? If μ assigns mass $\frac{1}{2}$ to each of the points $(\frac{1}{3}, \frac{1}{2})$ and $(\frac{2}{3}, \frac{1}{2})$?

Some of the ideas behind the proof of (5.1) will be brought out in the proof of the next theorem.

(5.2) THEOREM. *If K is a compact subset of S disjoint from the main diagonal, then each K-constructible distribution function is strictly singular.*

(5.3) DEFINITION. Suppose $\tau \in S^B$ has $\tau(b)$ in the interior of S for all $b \in B$. If n is a nonnegative integer, there are 2^n rectangles comprising $M_n(\tau)$; there are $2^n - 1$ exceptional points in I; for each other x in I, there is a unique one of these rectangles, call it the *n-th τ-rectangle over x*, whose projection on the horizontal axis contains x. This rectangle is $r(b, \tau)$ for a unique b in B_n; call it

$b(n, x, \tau)$. It is convenient to define these objects (somewhat arbitrarily) for all $\tau \in S^B$ and $x \in I$:

(i) $b(0, x, \tau) = \varnothing$;

(ii) $b(n+1, x, \tau) = b(n, x, \tau)0$ if $x \leq q(b(n, x, \tau), \tau_1)$;

(iii) $b(n+1, x, \tau) = b(n, x, \tau)1$ if $x > q(b(n, x, \tau), \tau_1)$.

Call $r(b(n, x, \tau), \tau)$ the n-th *τ-rectangle over x*.

The n-th τ-rectangle over x includes the $n+1$st; and its projection on the horizontal axis contains x. For each x and τ, $\{b(n, x, \tau): n \geq 0\}$ is a path through B.

A distribution function is *singular at a point* if it does not have a finite, positive derivative there. Theorem (5.2) is an immediate consequence of

(5.4) LEMMA. *If K is a compact subset of S disjoint from the main diagonal, $\tau \in T$, $x \in I$, and $\tau(b(j, x, \tau)) \in K$ for infinitely many nonnegative integers j, then $M(\tau)$ is singular at x.*

PROOF. Join each point of K by a line segment to $(0, 0)$ and by a line segment to $(1, 1)$. An easy compactness argument proves there is a $\delta > 0$ such that all the resulting line segments have slopes differing from 1 by δ or more. Suppose $M(\tau)$ is continuous at x, and $M(\tau)(x - \epsilon) < M(\tau)(x) < M(\tau)(x + \epsilon)$ for all $\epsilon > 0$; otherwise the conclusion is obvious. For each n, the n-th τ-rectangle over x has an interior. The diagonals of these rectangles form a sequence of chords inscribed in $M_\infty(\tau)$, whose projections on the horizontal axis shrink to x, such that infinitely often the ratio of the slopes of the successive chords differs from 1 by δ or more. ♦

The first step in proving (5.1) is to generalize (5.4), for which additional notation is needed.

(5.5) NOTATION. For $b \in B_i$ and $b' \in B_j$, let bb' be the element of B_{i+j} which agrees with b in its first i coordinates, and with b' in its last j. For $b \in B$ and $\tau \in S^B$ define $\tau[b] \in S^B$ by the relation

$$\tau[b](b') = \tau(bb') \quad \text{for all} \quad b' \in B. \tag{5.6}$$

Let

$$B(n) = \bigcup_{j=0}^{n} B_j. \tag{5.7}$$

If E is a subset of S and n is a nonnegative integer, let

$$E(n) = \{\tau: \tau \in S^B \text{ and } \tau(b) \in E \text{ for all } b \in B(n)\}, \tag{5.8}$$

and let

$$E_n = \{\tau: \tau \in S^B \text{ and } \tau(b) \in E \text{ for at least one } b \in B(n)\}. \tag{5.9}$$

(5.10) LEMMA. *If K is a compact subset of S disjoint from the main diagonal, C is a compact subset of the main diagonal disjoint from $(0, 0)$ and $(1, 1)$, $\tau \in T$, $x \in I$, n is a nonnegative integer, and $\tau[b(j, x, \tau)] \in (K \cup C)(n) \cap K_n$ for infinitely many nonnegative integers j, then $M(\tau)$ is singular at x.*

Lemma (5.4) is the special case: $n = 0$ and C is empty.

PROOF. There is a $\delta > 0$ with this property: if $\tau' \in (K \cup C)(n) \cap K_n$, and

k is the least nonnegative integer $\leq n$ with $\tau'(b') \in K$ for a $b' \in B_k$, the line segments joining $p(b', \tau')$ to $(0, 0)$ and $(1, 1)$ have slopes differing from 1 by δ or more. The rest of the proof is similar to that of (5.4). ♦

(5.11) NOTATION. For each subset A of S^B and nonnegative integers k and j, let:

(5.12) $\{A; k; j\} = \{\tau: \tau \in S^B$ and for each path $\{b_n: n \geq 0\}$ through B there are k or more n's with $\tau[b_n] \in A$ and $n \leq j\}$;

(5.13) $$\{A; k\} = \bigcup_j \{A; k; j\};$$

and

(5.14) $$\{A; \infty\} = \bigcap_k \{A; k\}.$$

If A is Borel in S^B, so is $\{A; k\}$.

(By König's lemma [15], $\{A; k\}$ is just the set of all $\tau \in S^B$ such that each path $\{b_n: n \geq 0\}$ through B has k or more n's with $\tau[b_n] \in A$. If this less cumbersome definition had been used, it would have been necessary to appeal to König's lemma in the proof of (5.18) below.)

Plainly, (5.1) is implied by (5.10) and (5.23). The balance of this section is occupied with proving (5.23). This will be easy with the help of (5.18), whose proof uses lemma (5.15).

(5.15) LEMMA. *Suppose* $0 < a < 1$ *and* $b \geq 1$ *with* $b(1 - a) < 1$. *If* θ *in* I *satisfies*

(5.16) $$\theta \geq a + (1 - a)\theta^b;$$

then $\theta = 1$.

PROOF. The function $x \to a + (1 - a)x^b$ is positive at $x = 0$, convex on I, and has derivative <1 at $x = 1$. Therefore, its graph lies wholly above the main diagonal of S on the half-open interval $[0, 1)$. ♦

(5.17) DEFINITION. For each subset B^* of B, a subset A of S^B is B^*-*dependent* if $\tau \in A$, $\tau' \in S^B$ and $\tau(b) = \tau'(b)$ for all $b \in B^*$ imply $\tau' \in A$. For example, the set $(K \cup C)(n) \cap K_n$ of (5.23) is $B(n)$-dependent.

(5.18) LEMMA. *If* n *is a nonnegative integer,* A *a* $B(n)$-*dependent Borel subset of* S^B, *and*

(5.19) $$b(n)(1 - \mu^B(A)) < 1,$$

where $b(n) = 2^{n+1}$, *then* $\{A; \infty\}$ *has* μ^B-*probability* 1.

PROOF. It is enough to prove that the μ^B-probability of $\{A; 1\}$, call it θ, is 1. Plainly,

(5.20) $$A \subset \{A; 1\},$$

and

(5.21) the conditional μ^B-probability of $\{A; 1\}$ given that A does not occur is at least $\theta^{b(n)}$.

In view of (5.20) and (5.21),

(5.22) $$\theta \geq \mu^B(A) + (1 - \mu^B(A))\theta^{b(n)},$$

and (5.15) applies. ♦

(The proof of (5.18) is a variant of a standard argument in the theory of branching processes ([10], XII.5). The lemma itself can be widely generalized.)

(5.23) LEMMA. *Unless μ assigns probability 1 to the main diagonal, there exists a compact subset C of the main diagonal containing neither $(0, 0)$ nor $(1, 1)$, a compact subset K of S disjoint from the main diagonal, and a nonnegative integer n, such that $\{(K \cup C)(n) \cap K_n; \infty\}$ has μ^B-probability 1.*

PROOF. Let D be all of S except the main diagonal. Since $\mu(D) > 0$, there is a positive integer n so large that

(5.24) $$b(n)(1 - \mu(D))^{b(n)-1} < 1.$$

Since μ assigns measure 0 to the corners $(0, 0)$ and $(1, 1)$ of S, there is a compact subset K of D, and a compact subset C of the main diagonal disjoint from $(0, 0)$ and $(1, 1)$, with $\mu(C) + \mu(K)$ so close to 1 that (5.24) implies

(5.25) $$\begin{aligned} b(n)[1 - \mu^B((K \cup C)(n) \cap K_n)] \\ &\leq b(n)\mu^B\{\tau: \tau \notin (K \cup C)(n)\} + b(n)\mu^B\{\tau: \tau \notin K_n\} \\ &= b(n)[1 - (\mu(K) + \mu(C))^{b(n)-1}] + b(n)[1 - \mu(K)]^{b(n)-1} \\ &< 1. \end{aligned}$$

Apply (5.18). ♦

6. When are all constructible distribution functions purely singular?

A subset K of S is *tangent* to the main diagonal if, for every strictly convex $F \in \Delta$ and strictly concave $G \in \Delta$, with $F(0) = G(0) = 0$, there is a point $(x, y) \in K$ with $F(x) < y < G(x)$.

(6.1) THEOREM. *Let K be a subset of S. If K is tangent to the main diagonal, then there is a K-constructible distribution function equivalent to Lebesgue measure; otherwise every K-constructible distribution function is purely singular.*

Preliminaries to the proof. Let 2 be the two-point set $\{0, 1\}$, and 2^∞ the space of infinite sequences of 0's and 1's. Let $\xi_1, \xi_2, \cdots$ be the coordinate process on 2^∞. Each $t \in I^B$ determines a unique probability $p(t)$ on 2^∞ by the relation: the $p(t)$-probability that $\xi_1 = 0$ is $t(\emptyset)$; the conditional $p(t)$ probability that $\xi_{n+1} = 0$ given $\xi_1, \cdots, \xi_n$ is $t(\xi_1 \cdots \xi_n)$. Let $\tau \in S^B$. These four facts are easily verified:

(6.2) τ is not in T if and only if there is a point in 2^∞ to which $p(\tau_1)$ and $p(\tau_2)$ both assign positive mass.

(6.3) If $\tau \in T$ and $F = M(\tau)$, then $F(x) = F_2(F_1^{-1}(x))$, where F_i is the $p(\tau_i)$-distribution function of $\sum_1^\infty \xi_n/2^n$ for $i = 1, 2$ (at least if F_1 and F_2 are continuous and strictly increasing).

(6.4) If $\tau \in T$ and $p(\tau_1)$ is equivalent to $p(\tau_2)$, then $M(\tau)$ is equivalent to Lebesgue measure.

(6.5) If $p(\tau_1)$ and $p(\tau_2)$ are mutually singular, then $M(\tau)$ is purely singular.

PROOF OF THE FIRST PART OF (6.1). Suppose $(x_n, y_n) \in K$, $0 < x_n$, $y_n < 1$, $x_n \to 0$, $x_n^{-1}y_n \to 1$. No new difficulty arises in other cases. By repeating terms, suppose

(6.6) $$\sum_0^\infty x_n = \infty.$$

By (4.13), suppose also

(6.7) $$\sum_0^\infty x_n\left(1 - \frac{y_n}{x_n}\right)^2 < \infty.$$

Define $\tau \in K^B$ by

(6.8) $$\tau(b) = (x_n, y_n) \quad \text{for all} \quad b \in B_n \quad \text{and} \quad n \geq 0.$$

Since $x_n \to 0$, (6.2) and (6.6) imply $\tau \in T$.

Under $p(\tau_1)$ (respectively, $p(\tau_2)$) the coordinate process is independent, and ξ_{n+1} is 0 with probability x_n (respectively, y_n). It now follows from (6.7), a calculation, and [13] that

(6.9) $$p(\tau_1) \text{ is equivalent to } p(\tau_2);$$

but a direct proof is easy and will be sketched here.

Let $\zeta_j = (1 - x_j)^{-1}(1 - y_j)\xi_{j+1} + x_j^{-1}y_j(1 - \xi_{j+1})$. By a familiar martingale argument, $(1 + \Pi_0^\infty \zeta_j)^{-1} \Pi_0^\infty \zeta_j$ is a version of the Radon-Nikodym derivative of $p(\tau_2)$ with respect to $(p(\tau_1) + p(\tau_2))$, so (6.9) follows from

(6.10) $$\Pi_0^\infty \zeta_j > 0 \text{ with } p(\tau_1)\text{-probability } 1,$$

and

(6.11) $$\Pi_0^\infty \zeta_j < \infty \text{ with } p(\tau_2)\text{-probability } 1,$$

which in turn follow from (6.7) and ([3], theorem 2.3, p. 108), because $\log \zeta_j$ has $p(\tau_i)$-mean essentially $\frac{1}{2}(-1)^i x_j(1 - x_j^{-1}y_j)^2$, and $p(\tau_i)$-mean-square essentially $x_j(1 - x_j^{-1}y_j)^2$, for $i = 1, 2$. In view of (6.4) and (6.9), the proof of the first part of (6.1) is complete.

PROOF OF THE SECOND PART OF (6.1). It will be shown that if $\tau \in K^B$, the continuous parts of $p(\tau_1)$ and $p(\tau_2)$ are mutually singular. The second part of (6.1) then follows from (6.2) and (6.5). There is an x with $0 < x < \frac{1}{2}$ such that K does not intersect the interior of the parallelogram whose vertices are $(0, 0)$, $(x, 1 - x)$, $(1, 1)$, and $(1 - x, x)$. Let G be the part of K included in the union of the two closed triangles with vertices $(0, 0)$, $(1 - x, x)$, $(1, 0)$, and $(0, 1)$, $(1, 1)$, $(x, 1 - x)$. (Several annoying difficulties in the rest of the argument disappear if K is the line segment joining $(0, 0)$ and $(1 - x, x)$.) Let H be $K - G$ (see figure 6.1). Let $\zeta_n = \xi_n$ if $\tau(\xi_1 \cdots \xi_{n-1})$ is above the cross diagonal, that is, the set of all $s \in S$ with $s(1) + s(2) = 1$; otherwise, $\zeta_n = 1 - \xi_n$. Let R be the (countable) set of $\omega \in 2^\infty$ such that $\zeta_n(\omega) = 0$ for all except finitely many n. It will be shown that $p(\tau_1)$ and $p(\tau_2)$ are mutually singular when restricted to $2^\infty - R$.

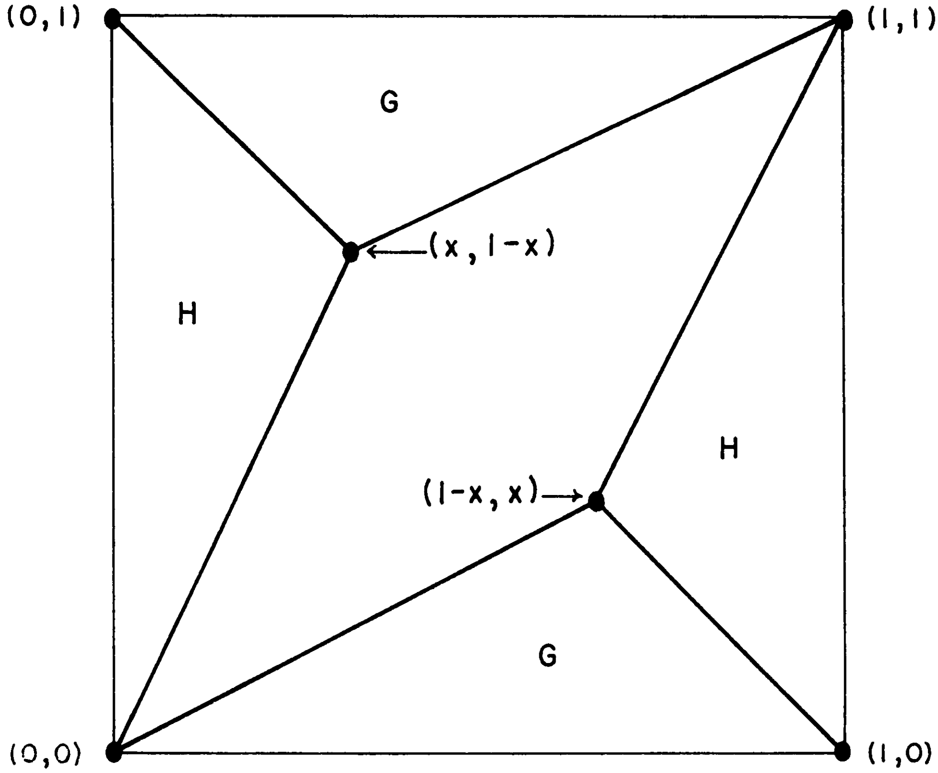


FIGURE 6.1

Indeed, $2^\infty - R$ is the union of $G_{1\infty}$ and $H_{1\infty}$, where $G_{1\infty}$ (respectively, $H_{1\infty}$) is the set of all $\omega \in 2^\infty$ such that $\tau[\xi_1(\omega) \cdots \xi_n(\omega)] \in G$ (respectively, H) and $\zeta_{n+1}(\omega) = 1$ for infinitely many n. On $G_{1\infty}$, let $\nu_j(\omega)$ be the j-th $n \geq 0$ with $\tau[\xi_1(\omega) \cdots \xi_n(\omega)] \in G$, for $j = 1, 2, \cdots$ and let $\eta_j = \zeta_{\nu_j+1}$. Off $G_{1\infty}$, let η_j be 0. Let $\mathfrak{F}_j$ be the σ-field generated by all the Borel subsets of $2^\infty - G_{1\infty}$, and by those Borel subsets of $G_{1\infty}$ which depend on $\xi_1, \cdots, \xi_{\nu_j+1}$ for $j = 0, 1, \cdots$. Let q_{ij} be the conditional $p(\tau_i)$-probability that $\eta_j = 1$, given $\mathfrak{F}_{j-1}$. By (1) of [6],

$$\lim_{n\to\infty} \frac{\eta_1 + \cdots + \eta_n}{q_{i1} + \cdots + q_{in}} = 1 \quad p(\tau_i)\text{-almost everywhere on } G_{1\infty}. \tag{6.12}$$

But

$$\frac{q_{11} + \cdots + q_{1n}}{q_{21} + \cdots + q_{2n}} \geq \frac{1-x}{x} > 1 \text{ on } G_{1\infty}, \tag{6.13}$$

so $p(\tau_1)$ and $p(\tau_2)$, when restricted to $G_{1\infty}$, are mutually singular. A similar argument applies to $H_{1\infty}$. ♦

7. When are all constructible distribution functions strictly singular?

There is a compact subset K of S such that every K-constructible distribution function is purely singular, yet some K-constructible distribution function is not strictly singular (7.4). We do not know a necessary and sufficient condition on K for every K-constructible distribution function to be strictly singular, but here is a sufficient condition.

(7.1) THEOREM. *Let K be a subset of S such that, for some strictly convex $F \in \Delta$ with $F(0) = 0$ every point of K is on or below the graph of F. Then there is a strictly convex $U \in \Delta$ with $U(0) = 0$ such that each K-constructible distribution function is everywhere less than or equal to U. Moreover, each K-constructible distribution function is strictly singular.*

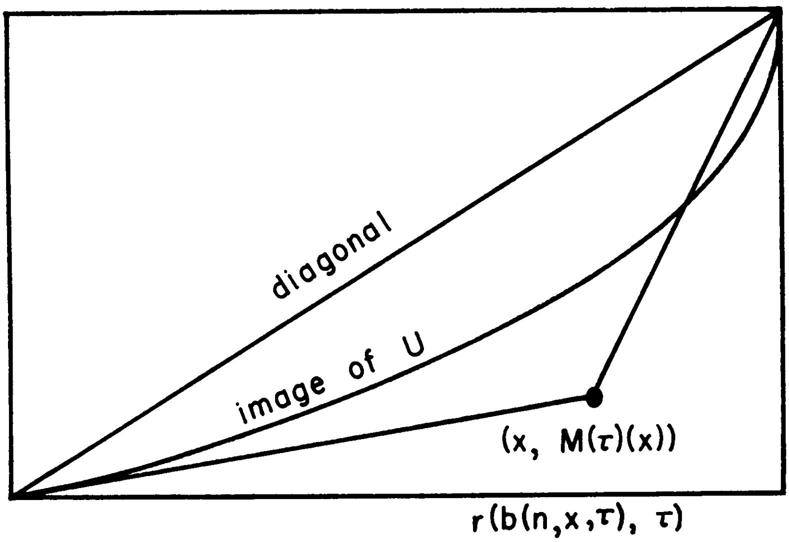


FIGURE 7.1

PROOF. Let t be the infimum of $1 - [(1-x)F(x)]/[x(1-F(x))]$ over all x with $0 < x < 1$. Since F is convex, $t > 0$. Let $U(x) = x(1-t)/(1-tx)$. Plainly, $U \in \Delta$ is strictly convex, and $F(x) \geq U(x)$ for all $x \in I$.

From ([8], theorems 9.2.2 and 4.2.1), if p and q are points on the graph of U, and S is mapped in a positive, affine way onto $p \times q$ (as defined in the paragraph following (3.20)), then the image of the graph of U is nowhere above the graph of U. Consequently, if p and q are points of S not above the graph of U, and S is mapped in a positive, affine way onto $p \times q$, the image of the graph of U is nowhere above the graph of U. This property of U, together with the fact that every point of K is on or below the graph of U, easily implies the first conclusion of the theorem.

Since U is strictly convex, there is an $\epsilon > 0$ such that: if $p \in S$ is neither $(0, 0)$ nor $(1, 1)$, and is not above the graph of U, then the slope of at least one of the two line segments joining p to $(0, 0)$ and to $(1, 1)$ differs from 1 by more than ϵ.

Let $\tau \in T \cap K^B$ and $x \in I$. Suppose $(x, M(\tau)(x))$ is in the interior of the n-th

τ-rectangle over x, $r(b(n, x, \tau), \tau)$, for all $n \geq 0$ (no new difficulty arises in other cases). If S is mapped in a positive, affine way onto $r(b(n, x, \tau), \tau)$, by the first conclusion of (7.1), the point $(x, M(\tau)(x))$ is not above the image of the graph of U. By the preceding paragraph, of the two line segments joining $(x, M(\tau)(x))$ to the lower left corner and to the upper right corner of $r(b(n, x, \tau), \tau)$, at least one has a slope whose ratio to the slope of the diagonal of $r(b(n, x, \tau), \tau)$ differs from 1 by more than ϵ (see figure 7.1). Consequently, the right and left derivatives of $M(\tau)$ at x cannot both exist and be equal to the same finite, positive number. ♦

We thank Georges De Rham for a proof of the interesting special case of the next theorem where K consists of a single point.

(7.2) THEOREM. *If K is a compact subset of S disjoint from the boundary and from the main diagonal, then no K-constructible distribution function has a finite, positive, one-sided derivative anywhere.*

PROOF. Let $\tau \in K^B$, and $0 \leq x < 1$. It will be argued that $M(\tau)$ has no finite, positive, right derivative at x. Suppose $(x, M(\tau)(x))$ is in the interior of the n-th τ-rectangle over x, for all $n \geq 0$ (no new difficulty arises in other cases). Let N be the set of $n = 0, 1, \cdots$ for which the point $p(b(n, x, \tau), \tau)$ is to the right of the vertical line through x. Plainly, N is infinite. By an easy compactness argument, there is an $\epsilon > 0$ such that if $n \in N$, and any point of $r(b(n, x, \tau)0, \tau)$ is joined by line segments (as in figure 7.2) to the upper right corner of $r(b(n, x, \tau), \tau)$, to $p(b(n, x, \tau), \tau)$, and to $p(b(n, x, \tau)1, \tau)$, then at least one pair of these three line segments have slopes whose ratio differs from 1 by more than ϵ. ♦

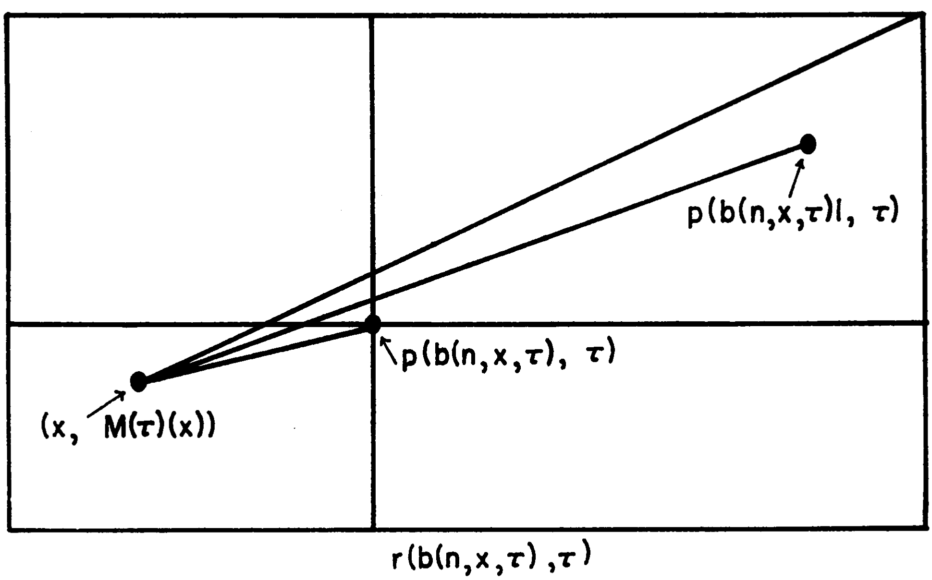


FIGURE 7.2

(7.3) EXAMPLE. Let K consist of the line segments joining $(\frac{2}{3}, \frac{1}{3})$ to $(0, 0)$ and to $(1, 1)$. Then K satisfies the hypotheses of (7.1), but not the conclusion of (7.2).

PROOF. Define $\tau_n \in S$ and $b_n \in B_n$: $b_0 = \varnothing$, $b_{2n+1} = b_{2n}1$, and $b_{2n+2} = b_{2n+1}0$; $\tau_{2n} = (2/(n + 10), 1/(n + 10))$, and $\tau_{2n+1} = (1 - 1/(n + 10), 1 - 2/(n + 10))$. If $\tau \in K^B \cap T$ has $\tau(b_n) = \tau_n$ for $n = 0, 1, \cdots$ and $x \in I$ has $b(n, x, \tau) = b_n$, $n = 0, 1, \cdots$, then $M(\tau)$ has finite, positive one-sided derivatives at x. Indeed, let δ_n be the width of $r(b(n, x, \tau), \tau)$. Then $\delta_0 = 1$, $\delta_{2n+1} = (1 - 2/(n + 10))\delta_{2n}$, and $\delta_{2n+2} = (1 - 1/(n + 10))\delta_{2n+1}$. Moreover, if the lower left corner of $r(b(n, x, \tau), \tau)$ is (x_n, y_n), then $x_0 = 0$, $x_{2n+1} = x_{2n} + 2\delta_{2n}/(n + 10)$, and $x_{2n+2} = x_{2n+1}$; similarly for y. Thus, one can estimate the chordal slope from (x_n, y_n) to $(x, M(\tau)(x))$. It converges to a finite, positive number λ. Since the ratio of $x - x_n$ to $x - x_{n+1}$ and of $M(\tau)(x) - y_n$ to $M(\tau)(x) - y_{n+1}$ converge to 1, λ is in fact the left derivative of $M(\tau)$ at x. Proceed similarly for the right derivative. ♦

(7.4) EXAMPLE. Let K consist of the line segments joining $(\frac{2}{3}, \frac{1}{3})$ to $(0, 0)$ and to $(1, 1)$, and of the line segments joining $(\frac{1}{3}, \frac{2}{3})$ to $(0, 0)$ and to $(1, 1)$. Then K satisfies the hypotheses of (6.1), but not the second conclusion of (7.1).

PROOF. Define $b_n \in B_n$ and $\tau_n \in S$: $b_0 = \varnothing$,

$$\begin{aligned}
\tau_{4n} &= (2/(n + 10), 1/(n + 10)), & b_{4n+1} &= b_{4n}1, \\
\tau_{4n+1} &= (1 - 1/(n + 10), 1 - 2/(n + 10)), & b_{4n+2} &= b_{4n+1}0, \\
\tau_{4n+2} &= (1/(n + 10), 2/(n + 10)), & b_{4n+3} &= b_{4n+2}1, \\
\tau_{4n+3} &= (1 - 2/(n + 10), 1 - 1/(n + 10)), & b_{4n+4} &= b_{4n+3}0.
\end{aligned} \tag{7.5}$$

If $\tau \in K^B \cap T$ has $\tau(b_n) = \tau_n$, and $x \in I$ has $b(n, x, \tau) = b_n$, for all n, then $M(\tau)$ has a finite, positive derivative at x. The reasoning for (7.3) applies. ♦

8. Mutual singularity of priors

If F and G are distribution functions, and there is no x for which the ratio of $F(x + h) - F(x)$ to $G(x + h) - G(x)$ converges to a finite, positive limit as $h \to 0$, then F is *strictly singular with respect to* G. If P and Q are probabilities on Δ, and there are Borel subsets C and D of Δ such that $P(C) = Q(D) = 1$ and every $F \in C$ is strictly singular with respect to every $G \in D$, then P is *strictly singular with respect to* Q.

(8.1) THEOREM. *Let $0 < r < 1$ and μ and ν be distinct base probabilities assigning measure 1 to the vertical line segment $\{s\colon s \in S, s(1) = r, 0 < s(2) < 1\}$. Then P_μ is strictly singular with respect to P_ν.*

In particular, $P_\mu \neq P_\nu$. We do not know when more general μ and ν lead to distinct P_μ and P_ν.

There is no real loss in setting r equal to $\frac{1}{2}$ in (8.1), the essential ideas of the proof already appearing when μ and ν assign measure 1 to the same two-point subset of the vertical segment $\{s\colon s \in S,\ s(1) = \frac{1}{2},\ 0 < s(2) < 1\}$ (case 3, below). The proof is easier if also μ is a one-point measure (case 2), and is easiest in case 1.

Throughout this section, let c and d be positive numbers less than 1 with $c \neq d$, and $\delta(z)$ the probability measure concentrated on the one-point set $\{z\}$.

Case 1. $\mu = \delta(\frac{1}{2}, c)$ and $\nu = p\delta(\frac{1}{2}, d) + (1 - p)\delta(\frac{1}{2}, c)$, with $\frac{1}{2} \leq p \leq 1$.

PROOF FOR CASE 1. Let $D^* = \{\tau\colon \tau \in S^B,\ b \in B$ implies $\tau(b)$ is $(\frac{1}{2}, c)$ or $(\frac{1}{2}, d)$, and for each path $b_0, b_1, \cdots$ through B, there are infinitely many n with $\tau(b_n) = (\frac{1}{2}, d)\}$, and let $D = M(D^*)$. As is clear from (5.18), or from ([10], XII.5), $\nu^B(D^*) = 1$, so $P_\nu(D) = 1$. Let $x \in I$ and $\tau \in D^*$. Let $\sigma(b) = (\frac{1}{2}, c)$ for $b \in B$, so $M(\sigma)$ is the coin-tossing distribution function Q_c. As will now be argued, the ratio of $M(\sigma)(x + h) - M(\sigma)(x)$ to $M(\tau)(x + h) - M(\tau)(x)$ does not converge to a finite, positive limit as $h \to 0$. The n-th σ-rectangle $r(b(n, x, \sigma), \sigma)$ over x (definition (5.3)) projects onto the same interval of the horizontal axis as does $r(b(n, x, \tau), \tau)$, namely, the leftmost of the intervals $[0, (1/2^n)]$, $[(1/2^n), (2/2^n)]$, $\cdots$, $[1 - (1/2^n), 1]$ which contains x. This interval shrinks to x as $n \to \infty$. What must be seen, therefore, is that the ratio r_n of the height of $r(b(n, x, \sigma), \sigma)$ to the height of $r(b(n, x, \tau), \tau)$ does not converge to a finite, positive limit as $n \to \infty$. Indeed, it is apparent that r_{n+1}/r_n does not converge to 1; for whenever $\tau(b(n, x, \tau)) = (\frac{1}{2}, d)$, r_{n+1}/r_n is c/d or $(1 - c)/(1 - d)$, according as the $(n + 1)$-st digit in the nonterminating binary expansion of x is 0 or 1. ♦

Case 2. $\mu = \delta(\frac{1}{2}, c)$ and $\nu = p\delta(\frac{1}{2}, d) + (1 - p)\delta(\frac{1}{2}, c)$, with $0 < p \leq 1$.

PROOF FOR CASE 2. Let E be the one-point set $\{(\frac{1}{2}, d)\}$. For every positive integer k, the subsets E_k and $(E_k; \infty)$ of S^B are defined by (5.9) and (5.14). By (5.18), there is a positive k so large that $(E_k; \infty)$ has probability 1 under ν^B, as does D^*, the intersection of $(E_k; \infty)$ with the set of $\tau \in S^B$ such that $\tau(b)$ is $(\frac{1}{2}, c)$ or $(\frac{1}{2}, d)$ for all $b \in B$. So the P_ν-probability of $M(D^*)$ is 1. The rest of the proof is similar to that for case 1, or is immediate from (8.2) below. ♦

Case 3. $\mu = q\delta(\frac{1}{2}, d) + (1 - q)\delta(\frac{1}{2}, c)$ and $\nu = p\delta(\frac{1}{2}, d) + (1 - p)\delta(\frac{1}{2}, c)$, with $p \neq q$ and $0 < p, q < 1$.

Case 3 uses (8.2) to (8.5), and this definition: $F \in \Delta$ is *strictly singular with respect to* $G \in \Delta$ *at* x if the ratio of $F(x + h) - F(x)$ to $G(x + h) - G(x)$ does not converge to a finite, positive limit as $h \to 0$.

The first lemma uses (5.3), (5.6), and (5.7). Of course, if $x \in I$, $\tau \in S^B$, and $\tau_1(b) = \frac{1}{2}$ for all $b \in B$, then $b(n, x, \tau)$ is simply the first n digits in the nonterminating binary expansion of x.

(8.2) LEMMA. *Let σ and τ be functions from B to the two-point set $\{(\frac{1}{2}, c), (\frac{1}{2}, d)\}$, let $x \in I$, and let k be a nonnegative integer. If for infinitely many n there is a $b' \in B(k)$ such that $\sigma[b(n, x, \sigma)]$ and $\tau[b(n, x, \tau)]$ differ at b', then $M(\sigma)$ is strictly singular with respect to $M(\tau)$ at x.*

PROOF. The proof is easy. ♦

For real numbers p and α with $0 \leq p \leq 1$ and $0 < \alpha < 1$, let

$$m(p, \alpha) = \left(\frac{p}{\alpha}\right)^\alpha \left(\frac{1 - p}{1 - \alpha}\right)^{1-\alpha}. \tag{8.3}$$

An interesting fact in [1] is recorded here as

(8.4) LEMMA. *Let $X_1, X_2, \cdots$ be independent random variables, each assuming the value 1 with probability p, and 0 with probability $1 - p$. Let $p < \alpha < 1$, and let n be a positive integer. Then $X_1 + \cdots + X_n \geq n\alpha$ with probability no more than $[m(p, \alpha)]^n$.*

For probabilities P and Q on a σ-field $\mathfrak{F}$ of subsets of a set Ω, let P^n and Q^n be the power probabilities on Ω^n, as defined in section 2, and let $\|P - Q\| = \sup\{P(A) - Q(A): A \in \mathfrak{F}\}$.

(8.5) LEMMA. *For any pair P and Q of distinct probabilities on a measurable space $(\Omega, \mathfrak{F})$, there is a positive $\rho < 1$ such that for all $n \geq 1$, $\|P^n - Q^n\| > 1 - \rho^n$.*

PROOF. Let $A \in \mathfrak{F}$ and α satisfy $P(A) < \alpha < Q(A)$. For $i = 1, \cdots, n$ and $\omega \in \Omega^n$, let $X_i(\omega)$ be 1 or 0, according as the i-th coordinate of ω is or is not in A. As (8.4) implies

$$P^n(X_1 + \cdots + X_n \geq n\alpha) \leq [m(Q(A), \alpha)]^n, \tag{8.6}$$

and

$$Q^n(X_1 + \cdots + X_n \geq n\alpha) \geq 1 - [m(Q(A), \alpha)]^n. \tag{8.7}$$

◆

Let X_b be independent, identically distributed random variables, $X_b = 1$ with probability p, $X_b = 0$ with probability $1 - p$, where the index b ranges over B, the set of all finite sequences of 0's and 1's, including the empty sequence $\varnothing$. Of course, each infinite sequence of 0's and 1's, $y = (y_1, y_2, \cdots)$, determines the sequence of n-tuples of 0's and 1's, $b_0(y), b_1(y), \cdots$, where $b_n(y) = (y_1, \cdots, y_n)$. The strong law of large numbers plainly implies that for every y, the sequence of random variables $Z_n(y) = X_{b_n(y)}$ satisfies $(Z_1(y) + \cdots + Z_n(y))/n \to p$, except for an event N_y of probability 0. The null event N_y cannot be independent of y. Indeed, with probability 1, there are y for which $\limsup (Z_1(y) + \cdots + Z_n(y))/n$ exceeds p. However, for $p < \frac{1}{2}$, there is an α, $p < \alpha < 1$, such that with probability 1, for every y, $\limsup (Z_1(y) + \cdots + Z_n(y))/n$ is no more than α. This fact will be proved in a sharper and more general form (8.8), although we were unable to find the best α. Incidentally, if $p \geq \frac{1}{2}$, no such α exists.

Preliminaries to (8.8). Let j be a positive integer. For each nonnegative integer n, let J_n be the set of all n-tuples formed with the j integers $0, \cdots, j - 1$. The only element of J_0 is the (empty) 0-tuple $\varnothing$. For $b \in J_n$, and $i = 0, \cdots, j - 1$, b followed by i, namely bi, is in J_{n+1}. Let $J = \bigcup_{n=0}^{\infty} J_n$. A *path through* J is a sequence $b_0, b_1, \cdots$ such that $b_0 = \varnothing$ and for all n, there is an $i = 0, \cdots, j - 1$ with $b_{n+1} = b_n i$. Let P be the probability on the two-point set $\{0, 1\}$ which assigns probability p to 1. The power probability P^J on the set of functions from J to $\{0, 1\}$ was defined in section 2 and $m(p, \alpha)$ in (8.3).

(8.8) LEMMA. *If $0 \leq p < \alpha < 1$, and $m(p, \alpha) < j^{-1}$, then for P^J-almost all functions f from J to $\{0, 1\}$, there is an $n(f) < \infty$ such that: for each $n \geq n(f)$ and path $b_0, b_1, \ldots$ through J, $f(b_0) + \cdots + f(b_{n-1}) < n\alpha$.*

PROOF. Let E_n be the set of all functions g from J to $\{0, 1\}$ such that, for some path $b_0, b_1, \cdots$ through J, $g(b_0) + \cdots + g(b_{n-1}) \geq n\alpha$. By (8.4), $P^J(E_n) \leq j^{n-1}[m(p, \alpha)]^n$, which is summable in n. ♦

Proof for case 3. By (8.5), there is a positive $\rho < 1$ with $\|\mu^n - \nu^n\| > 1 - \rho^n$ for all positive integers n. Let

$$\rho_k = \rho^{2^{k+1}-1}. \tag{8.9}$$

For each nonnegative integer k, there is a $B(k)$-dependent (as defined by (5.7) and (5.17)) subset A of S^B such that:

(i) if $\tau \in A$ and $b \in B(k)$, then $\tau(b)$ is $(\frac{1}{2}, c)$ or $(\frac{1}{2}, d)$;

(ii) $\mu^B(A) > 1 - \rho_k$;

and

(iii) $\nu^B(A) < \rho_k$.

Choose k so large that $2^{k+1}m(\rho_k, \frac{1}{3}) < 1$.

Let C^* be the set of all $\tau \in S^B$ for which:

(i) $\tau(b)$ is $(\frac{1}{2}, c)$ or $(\frac{1}{2}, d)$ for all $b \in B$;

and

(ii) there is a positive integer $n(\tau)$ such that, for each $n \geq n(\tau)$ and path $b_0, b_1, \cdots$ through B, the number of nonnegative integers $i \leq n - 1$ with $\tau[b_{i(k+1)}] \in A$ is greater than $2n/3$.

Let D^* be the set of all $\tau \in S^B$ for which:

(i) $\tau(b)$ is $(\frac{1}{2}, c)$ or $(\frac{1}{2}, d)$ for all $b \in B$;

and

(ii) there is a positive integer $n(\tau)$ such that, for each $n \geq n(\tau)$ and path $b_0, b_1, \cdots$ through B, the number of nonnegative integers $i \leq n - 1$ with $\tau[b_{i(k+1)}] \in A$ is less than $n/3$.

Use (8.8), with $j = 2^{k+1}$, $\alpha = \frac{1}{3}$, $p = 1 - \mu^B(A)$ and $p = \nu^B(A)$, to see that $\mu^B(C^*) = \nu^B(D^*) = 1$. By (8.2), $\sigma \in C^*$ and $\tau \in D^*$ implies that $M(\sigma)$ is strictly singular with respect to $M(\tau)$. ♦

The proof for case 3 can easily be transformed into a proof of the full (8.1), especially with the aid of the following lemma, which though cumbersome to state, is easy to prove.

(8.10) Lemma. *Let $x \in I$ and k be a nonnegative integer. Let V and W be disjoint, compact, $B(k)$-dependent subsets of S^B, such that $\tau \in V \cup W$ and $b \in B(k)$ implies $\tau_1(b) = \frac{1}{2}$ and $0 < \tau_2(b) < 1$. Let $\sigma \in S^B$ and $\tau \in S^B$, with $\sigma_1(b) = \tau_1(b) = \frac{1}{2}$ and $0 < \sigma_2(b), \tau_2(b) < 1$ for all $b \in B$. If $\sigma[b(n, x, \sigma)] \in V$ and $\tau[b(n, x, \tau)] \in W$ for infinitely many n, then $M(\sigma)$ is strictly singular with respect to $M(\tau)$ at x.*

One of us (Freedman, *Ann. Math. Statist.*, Vol. 37 (1966), pp. 375–381) has extended (8.1) so as to permit μ and ν to assign positive measure to the two points $(r, 0)$ and $(r, 1)$.

9. The average distribution function

A probability P on Δ determines an *average distribution function* $F_P \in \Delta$, namely,

$$F_P(z) = \int_{G \in \Delta} G(z) P(dG). \tag{9.1}$$

This section studies F_{P_μ}, or F_μ for short, where μ is a base probability. We do not know which $F \in \Delta$ are of the form F_μ for some base probability μ.

(9.2) DEFINITION. If μ is a probability on S, then T_μ is this mapping of Δ into itself. If $G \in \Delta$, then $T_\mu G$ is the distribution function of a point v chosen from I according to this mechanism: choose a point u at random from I according to G, and independently a point (x, y) at random from S according to μ; then v is xu with probability y, and $u + x(1 - u)$ with probability $1 - y$.

For $0 \leq z < 1$,

$$(T_\mu G)(z) = \int_{0 \leq x \leq z, 0 \leq y \leq 1} \left[y + (1 - y)G\left(\frac{z - x}{1 - x}\right)\right] \mu(dx, dy) + \int_{z < x \leq 1, 0 \leq y \leq 1} yG\left(\frac{z}{x}\right) \mu(dx, dy); \tag{9.3}$$

and $(T_\mu G)(1) = 1$.

(9.4) DEFINITION. If μ is a probability on S, then L_μ, a probability on the linear functions from I to I, is the distribution of a linear function chosen according to this mechanism: choose (x, y) at random from S according to μ; then choose the function $u \to xu$ with probability y, and the function $u \to u + x(1 - u)$ with probability $1 - y$.

(9.5) DEFINITION. Let μ_1 be that subprobability on I for which $\mu_1[0, x]$ is the L_μ-probability of the set of linear functions for which $f(0) = 0$ and $f(1) \leq x$. Similarly, $\mu_0[0, x]$ is the L_μ-probability that $f(1) = 1$ and $f(0) \leq x$.

(9.6) DEFINITION. If P is a probability on the linear functions from I to I, then, in conformity with ([7], section 5), P^* is this mapping of Δ into itself. If $G \in \Delta$, then P^*G is the distribution function of $f(u)$, where f and u are chosen independently, the linear function f according to P and the point u according to G.

(9.7) DEFINITION. If μ is a probability on S, then μ_h is the projection of μ on the horizontal axis, and $\theta(\mu, x)$ is the conditional μ-expectation of $s \to s(2)$, given $s(1) = x$.

(9.8) LEMMA. *For probabilities μ and ν on S, the following conditions are equivalent:*

(i) $\mu_h = \nu_h$ *and* $\theta(\mu, \cdot) = \theta(\nu, \cdot)$;

(ii) $T_\mu = T_\nu$;

(iii) $L_\mu = L_\nu$;

(iv) $\mu_0 = \nu_0$ *and* $\mu_1 = \nu_1$.

PROOF. *Condition* (i) *implies* (ii). If $G \in \Delta$ and $0 \leq z < 1$,

$$(T_\mu G)(z) = \int_{0 \leq x \leq z} \theta(\mu, x)\mu_h(dx) + \int_{0 \leq x \leq z} [1 - \theta(\mu, x)]G\left(\frac{z - x}{1 - x}\right) \mu_h(dx) + \int_{z < x \leq 1} \theta(\mu, x)G\left(\frac{z}{x}\right) \mu_h(dx). \tag{9.9}$$

Condition (ii) *implies* (iii). Apply ([7], (6.2)) and the identity

$$T_\mu = (L_\mu)^*. \tag{9.10}$$

Condition (iii) *implies* (iv). This is clear.

Condition (iv) *implies* (i). Verify that $\mu_h = \mu_0 + \mu_1$, and that $\theta(\mu, x)$ is the Radon-Nikodym derivative of μ_1 with respect to μ_h at x; that is,

$$\theta(\mu, x) = \frac{\mu_1(dx)}{\mu_h(dx)}. \tag{9.11}$$ ♦

From now on, as usual, μ is a base probability.

(9.12) THEOREM. *F_μ is the unique fixed point of T_μ, and $F \in \Delta$ implies $(T_\mu)^n F \to F_\mu$.*

PROOF. If $0 \leq z \leq 1$, then $F_\mu(z)$ is the μ^B-expectation of the function $\tau \to M(\tau)(z)$. If $z < 1$, the conditional μ^B-expectation of $M(\cdot)(z)$ given $\tau(\varnothing) = (x, y)$ is

$$y + (1 - y)F_\mu\left(\frac{z - x}{1 - x}\right), \quad \text{provided} \quad x \leq z; \tag{9.13}$$

and

$$yF_\mu\left(\frac{z}{x}\right), \quad \text{provided} \quad x > z. \tag{9.14}$$

Integrating with respect to $\mu(dx, dy)$ proves $T_\mu F_\mu = F_\mu$. Apply (9.10) and ([7], (4.4)). ♦

(9.15) COROLLARY. *If μ and ν are base probabilities with $\mu_h = \nu_h$ and $\theta(\mu, \cdot) = \theta(\nu, \cdot)$, then $F_\mu = F_\nu$.*

PROOF. Use (9.12) and the relation (i) implies (ii) in (9.8). ♦

We guess that unless $F_\mu(z) = z$ for all $z \in I$, or F_μ assigns measure 0 to the interior of I, F_μ determines μ_h and $\theta(\mu, \cdot)$.

(9.16) DEFINITION. In conformity with the notation in ([8], chapters 5 and 6), for $0 < w < 1$, let $Q_w \in \Delta$ be the distribution function of $\sum_1^\infty X_i/2^i$, the X_i being independent, 0 with probability w, and 1 with probability $1 - w$; for $0 < r < 1$, let $S_{w,r} \in \Delta$ be $Q_w(Q_r^{-1})$.

(9.17) THEOREM. *If (r, w) is an interior point of S, μ assigns measure 1 to the vertical line segment $\{s\colon s \in S,\ s(1) = r\}$ and has mean (r, w), then $F_\mu = S_{w,r}$.*

PROOF. Apply (9.15). ♦

(9.18) LEMMA. *If $\theta(\mu, \cdot)$ has k continuous derivatives on I, and μ_h has a density with k continuous derivatives on I, then F_μ has k continuous derivatives on the interior of I.*

PROOF. From (9.9), for $0 < x < 1$, with m for the density of μ_h:

$$\begin{aligned} F_\mu(z) = {} & \int_0^z \theta(\mu, x)m(x)\,dx \\ & + \int_0^z \left[1 - \theta\left(\mu, \frac{z - v}{1 - v}\right)\right] m\left(\frac{z - v}{1 - v}\right) \frac{1 - z}{(1 - v)^2} F_\mu(v)\,dv \\ & + \int_z^1 \theta\left(\mu, \frac{z}{v}\right) m\left(\frac{z}{v}\right) \frac{z}{v^2} F_\mu(v)\,dv. \end{aligned} \tag{9.19}$$

Inductively on k, differentiate with respect to z. ♦

(9.20) THEOREM. *F_μ is continuous if and only if μ assigns measure 0 to the vertical edges of S, and measure less than 1 to each of the horizontal edges.*

PROOF. Apply ([7], (4.5) or (6.1)), using (9.10) and (9.12). ♦

(9.21) THEOREM. *If $0 < w < 1$, and μ is the uniform probability on the horizontal line segment $\{s: s \in S, s(2) = w\}$, then F_μ is absolutely continuous, with density on $(0, 1)$ proportional to*

$$z \to \frac{1}{z^w(1-z)^{1-w}}. \tag{9.22}$$

PROOF. By (9.19),

$$F_\mu(z) = wz + (1-w)\int_0^z \frac{1-z}{(1-v)^2} F_\mu(v)\,dv + w\int_z^1 \frac{z}{v^2} F_\mu(v)\,dv. \tag{9.23}$$

By (9.18), F_μ is absolutely continuous on $(0, 1)$, with infinitely differentiable density f_μ, and by (9.20), F_μ is continuous; so F_μ is absolutely continuous on I. Differentiating (9.23) twice with respect to z gives

$$f'_\mu(z) = f_\mu(z)\left(\frac{1-w}{1-z} - \frac{w}{z}\right) \quad \text{for} \quad 0 < z < 1. \tag{9.24}$$ ♦

(9.25) THEOREM. *If μ assigns measure 0 to the vertical edges of S, then F_μ is either purely singular or absolutely continuous.*

PROOF. In view of (9.10) and (9.12), ([7], (2.5)) applies. ♦

(9.26) LEMMA. *If μ_h is not purely singular, and $F \in \Delta$ assigns positive measure to the interior of I, then $T_\mu F$ is not purely singular.*

PROOF. Let F_z be the distribution function assigning measure 1 to z. Since $T_\mu F = \int (T_\mu F_z)F(dz)$, it is enough to check the special case, $F = F_z$, $0 < z < 1$. As (9.10) plainly implies, $T_\mu F_z$ restricted to $[0, z]$ is an affine image of μ_1; similarly for $[z, 1]$. So, if $\mu_h = \mu_0 + \mu_1$ is not purely singular, neither is $T_\mu F_z$. ♦

(9.27) THEOREM. *If μ assigns measure 0 to the vertical edges of S, measure less than 1 to each of the horizontal edges, and μ_h is not purely singular, then F_μ is absolutely continuous.*

PROOF. By (9.20), F_μ is continuous, so it assigns measure 1 to the interior of I, and $F_\mu = T_\mu F_\mu$ is not purely singular by (9.26). But F_μ is pure by (9.25). ♦

(9.28) THEOREM. *F_μ is the uniform distribution if and only if $\theta(\mu, x) = x$ for μ_h-almost all x.*

PROOF. For "if," suppose without loss of generality by (9.15) that μ assigns probability 1 to the main diagonal.

For "only if," by (9.10), the image of Lebesgue measure under T_μ has density at z equal to:

$$\int_{[0,z]} \frac{1}{1-x}\,\mu_0(dx) + \int_{(z,1]} \frac{1}{x}\,\mu_1(dx). \tag{9.29}$$

Since Lebesgue measure is fixed under T_μ, (9.29) is 1 for Lebesgue almost all $z \in I$. Since (9.29) is continuous from the right, it is 1 for all z with $0 \le z < 1$. By (9.20), μ_h assigns measure 0 to 0 and to 1. Therefore, setting $z = 0$ in (9.29),

$$\int_{(0,1]} \frac{1}{x}\,\mu_1(dx) = 1; \tag{9.30}$$

so

$$\int_{[0,z]} \frac{1}{1-x}\,\mu_0(dx) = \int_{(0,z]} \frac{1}{x}\,\mu_1(dx), \tag{9.31}$$

that is,

$$\frac{1}{1-x}\,\mu_0(dx) = \frac{1}{x}\,\mu_1(dx). \tag{9.32}$$

Consequently,

$$\frac{\mu_1(dx)}{\mu_h(dx)} = \frac{\mu_1(dx)}{\mu_0(dx) + \mu_1(dx)} = x. \tag{9.33}$$

Apply (9.11). ♦

(9.34) THEOREM. *If μ assigns positive measure to the interior of S, then F_μ is strictly increasing. More generally, F_μ is strictly increasing if and only if there are points x_0 and x_1 in the supports of μ_0 and μ_1 respectively with $x_0 \le x_1$.*

PROOF. The proof is easy, for example with the help of ([7], (5.17) and (6.1)). ♦

(9.35) DEFINITION. A mapping T of Δ into itself is a *uniformly strict contraction* if there is a nonnegative $\lambda < 1$, with

$$\sup_{z\in I} |(TF)(z) - (TG)(z)| \le \lambda \sup_{z \in I} |F(z) - G(z)|, \tag{9.36}$$

for all $F \in \Delta$ and $G \in \Delta$.

(9.37) THEOREM. *If μ assigns positive measure to the interior of S or to a vertical edge of S, then T_μ is a uniformly strict contraction of Δ.*

If μ assigns measure 0 to the vertical edges of S and the interior of S, then

(i) *T_μ is a uniformly strict contraction of Δ if and only if for some x with $0 < x < 1$, $\mu_0[x, 1] > 0$ and $\mu_1[0, x] > 0$;*

and

(ii) *some power of T_μ is a uniformly strict contraction of Δ if and only if μ assigns positive measure to each of the horizontal edges of S.*

PROOF. The result ([7], (5.10)) applies. ♦

10. Index of definitions

REFERENCES

[1] HERMAN CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 493–507.

[2] GEORGES DE RHAM, "Sur quelques courbes definies par des équations fonctionelles," *Rendiconti del Seminario Matematico dell'Università e del Politecnico di Torino*, Vol. 16 (1956–1957), pp. 101–113.

[3] J. L. DOOB, *Stochastic Processes*, New York, Wiley, 1953.

[4] LESTER E. DUBINS and DAVID A. FREEDMAN, "Random distribution functions," *Bull. Amer. Math. Soc.*, Vol. 69 (1963), pp. 548–551.

[5] ———, "Measurable sets of measures," *Pacific J. Math.*, Vol. 14 (1964), pp. 1211–1222.

[6] ———, "A sharper form of the Borel Cantelli lemma and the strong law," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 800–807.

[7] ———, "Invariant probabilities for certain Markov processes," *Ann. Math. Statist.*, Vol. 37 (1966), pp. 837–848.

[8] LESTER E. DUBINS and LEONARD J. SAVAGE, *How to Gamble If You Must*, New York, McGraw-Hill, 1965.

[9] J. FABIUS, "Asymptotic behavior of Bayes' estimates," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 846–856.

[10] WILLIAM FELLER, *An Introduction to Probability Theory and Its Applications*, Vol. 1, New York, Wiley, 1958 (2d ed.).

[11] DAVID A. FREEDMAN, "On the asymptotic behavior of Bayes' estimates in the discrete case," *Ann. Math. Statist.*, Vol. 34 (1963), pp. 1386–1403.

[12] F. HAUSDORFF, *Set Theory*, New York, Chelsea, 1957.

[13] SHIZUO KAKUTANI, "On the equivalence of infinite product sequences," *Ann. of Math.*, Vol. 49 (1948), pp. 214–224.

[14] J. R. KINNEY and T. S. PITCHER, "The dimension of the support of a random distribution function," *Bull. Amer. Math. Soc.*, Vol. 70 (1964), pp. 161–164.

[15] DENES KÖNIG, *Theorie der Endlichen und Unendlichen Graphen*, Leipzig, Akademische Verlags-Gesellschaft, 1936.

[16] CHARLES H. KRAFT, "A class of distribution function processes which have derivatives," *J. Appl. Probability*, Vol. 1 (1964), pp. 385–388.

[17] CHARLES H. KRAFT and CONSTANCE VAN EEDEN, "Bayesian bio-assay," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 886–890.

[18] CASIMIR KURATOWSKI, "Les fonctions semi-continues dans l'espace des ensembles fermés," *Fund. Math.*, Vol. 18 (1932), pp. 148–159.

[19] ———, *Topologie*, Vol. 1, New York Hafner, 1958 (4th ed.).

[20] ———, *Topologie*, Vol. 2, New York, Hafner, 1952 (3d ed.).