PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 92, Number 2, October 1984

FACTORIZATION OF PROBABILITY MEASURES AND ABSOLUTELY MEASURABLE SETS

DAVID BLACKWELL AND ASHOK MAITRA

ABSTRACT. We find necessary and sufficient conditions for a separable metric space Y to possess the property that for any measurable space $(X, \mathcal{A})$ and probability measure P on $X \times Y$, P can be factored.

1. Introduction. We characterize separable metric spaces Y which have the following property: for any measurable space $(X, \mathcal{A})$ and any probability measure P on the product space $(X \times Y, \mathcal{A} \times \mathcal{B}(Y))$, where $\mathcal{B}(Z)$ will denote the Borel σ-field of a metric space Z, P can be *factored*: $P = \mathcal{Q} \times T$, where $\mathcal{Q}$ is a probability measure on $(X, \mathcal{A})$ and T: $X \times \mathcal{B}(Y) \to [0,1]$ is an $\mathcal{A}$-measurable transition function such that

$$P(A \times B) = \int_A T(B|x)\, dQ(x)$$

for every $A \in \mathcal{A}$ and $B \in \mathcal{B}(Y)$.

The factorability of P is of obvious interest to the Bayesian statistician. For, imagine that $(Y, \mathcal{B}(Y))$ is the parameter space and that $(X, \mathcal{A})$ is the sample space. A prior on $(Y, \mathcal{B}(Y))$ together with a model (that is, a $\mathcal{B}(Y)$-measurable transition function T: $Y \times \mathcal{A} \to [0,1]$) determine a probability measure P on $(X \times Y,\ \mathcal{A} \times \mathcal{B}(Y))$. The Bayesian bases his inference on the posterior distribution of the parameter given the sample, so he needs to factor P as above.

The main result of the paper is as follows.

THEOREM. *Let Y be a separable metric space. Then the following conditions on Y are equivalent.*

(a) *Y is absolutely measurable, i.e., if $\tilde{Y}$ is a metric completion of Y and λ is a probability measure on the Borel σ-field of $\tilde{Y}$, then Y is λ-measurable.*

(b) *For any measurable space $(X, \mathcal{A})$ and any probability measure P on $(X \times Y, \mathcal{A} \times \mathcal{B}(Y))$, P can be factored.*

(c) *For any Polish space X and any probability measure P on $(X \times Y, \mathcal{B}(X) \times \mathcal{B}(Y))$, P can be factored.*

The above result gives a new characterization of absolutely measurable sets. Using the theory of compact measures developed by Marczewski [**3**], one can deduce the implication (a)$\to$(b) from a result of Jirina [**1**] and the implication (b)$\to$(a) from a result of Pachl [**4**]. The implication (c)$\to$(a) is new. Our proofs, however, are direct and elementary in nature and do not use the somewhat elaborate machinery of compact measures.

The proofs will be given in §3. §2 explains the notation to be used.

Received by the editors July 5, 1983 and, in revised form, October 24, 1983.
1980 *Mathematics Subject Classification*. Primary 28A05, 60A10, 60A05; Secondary 28A50.
Key words and phrases. Probability measure, product space, transition function, factorization, absolutely measurable set, prior distribution, posterior distribution.

2. Notation. If $(X, \mathcal{A})$ is a measurable space, $M(X, \mathcal{A})$ will denote the set of probability measures on $(X, \mathcal{A})$. If $(Z, \mathcal{C})$ is another measurable space, a function T: $X \times \mathcal{C} \to [0,1]$ is said to be a $(X, \mathcal{A}) - (Z, \mathcal{C})$ *transition function* if, for each $x \in X$, $T(\cdot|x)$ is a probability measure on $(Z, \mathcal{C})$, and for each $C \in \mathcal{C}$, $T(C|\cdot)$ is an $\mathcal{A}$-measurable function on X. The set of $(X, \mathcal{A})-(Z, \mathcal{C})$ transition functions is denoted by $M(Z, \mathcal{C}|X, \mathcal{A})$. If X, Z are metric spaces, we write $M(X)$ for $M(X, \mathcal{B}(X))$ and $M(Z|X)$ for $M(Z, \mathcal{B}(Z)|X, \mathcal{B}(X))$.

We denote by 2^N the set of infinite sequences of 0's and 1's and by $\mathcal{S}$ the set of finite sequences of 0's and 1's of positive length. For each $\mathfrak{s} \in \mathcal{S}$, $L(\mathfrak{s})$ will denote the set of elements of 2^N that begin with $\mathfrak{s}$. The sets $L(\mathfrak{s})$, $\mathfrak{s} \in \mathcal{S}$, form a base for a topology on 2^N. Endowed with this topology, 2^N is a homeomorph of the Cantor ternary set.

If Z is a metric space and $\lambda \in M(Z)$, λ^* and λ_* denote, respectively, the outer measure and inner measure induced by λ.

3. Proofs. We begin with two lemmas, both of which are well known, and only the second will be proved here.

LEMMA 1. *Suppose* f: $\mathcal{S} \to [0,1]$ *satisfies* (i) $f(0) + f(1) = 1$, *and* (ii) $f(\mathfrak{s}0) + f(\mathfrak{s}1) = f(\mathfrak{s})$ *for each* $\mathfrak{s} \in \mathcal{S}$. *Then there is a unique* $\mu \in M(2^N)$ *such that* $\mu(L(\mathfrak{s})) = f(\mathfrak{s})$ *for each* $\mathfrak{s} \in \mathcal{S}$.

LEMMA 2. *Suppose X is a Polish space and let $V_\mathfrak{n}$, $\mathfrak{n} \geq 1$, be a base for the topology of X. Let φ: $X \to 2^N$ be defined by: $\varphi(x) = (I_{V_1}(x), I_{V_2}(x), \ldots)$, where I_{V_i} denotes the indicator function of the set V_i. Then φ is a Borel isomorphism of X and $\varphi(X)$, and $\varphi(X)$ is Borel in 2^N (indeed, a G_δ in 2^N).*

PROOF. Since the sets $V_\mathfrak{n}$ separate points of X, the function φ is one-one. It is easy to see that the inverse image of an open set under φ is a F_σ set in X and the forward image of an open set under φ is open in $\varphi(X)$. In other words, φ^{-1} is continuous. So by a well-known result [**2**, Corollary 3, p. 436], there is a G_δ subset G of 2^N and a continuous function $\mathfrak{g}$: $G \to X$ such that $\varphi(X) \subseteq G$ and $\mathfrak{g} = \varphi^{-1}$ on $\varphi(X)$. Consequently, $\varphi(X) = \{\mathfrak{z} \in G: \varphi(\mathfrak{g}(\mathfrak{z})) = \mathfrak{z}\}$. Hence, if $W_\mathfrak{n}$, $\mathfrak{n} \geq 1$, is a base for the topology of G, then

$$G \setminus \varphi(X) = \bigcup_{\mathfrak{n} \geq 1} [\mathfrak{g}^{-1}(\varphi^{-1}(W_\mathfrak{n})) \cap (G \setminus W_\mathfrak{n})],$$

so that $G \setminus \varphi(X)$ is a F_σ set in G. It follows that $\varphi(X)$ is a G_δ in G, so a G_δ in 2^N. The second sentence of the proof already establishes that φ is a Borel isomorphism of X and $\varphi(X)$.

We now turn to the proof of the Theorem. The implication (a)$\to$(b) is proved first for Y a Borel subset of 2^N, then for Y a Borel subset of a Polish space, and finally for Y an absolutely measurable set.

Let, then, Y be a Borel subset of 2^N. Suppose P is a probability measure on $(X \times Y, \mathcal{A} \times \mathcal{B}(Y))$. We define P' on $(\mathcal{X} \times 2^N, \mathcal{A} \times \mathcal{B}(2^N))$ by setting $P'(E) = P(E \cap (X \times Y))$ for $E \in \mathcal{A} \times \mathcal{B}(2^N)$, so P' is a probability measure. For each $\mathfrak{s} \in \mathcal{S}$, fix a version $P'(L(\mathfrak{s})|\mathcal{A}^*)$ of the conditional probability under P' of the set $X \times L(\mathfrak{s})$ given $\mathcal{A}^*$, where $\mathcal{A}^*$ is the σ-field on $X \times 2^N$ of events of the form $A \times 2^N$ with $A \in \mathcal{A}$. For each $x \in \mathcal{X}$ and $\mathfrak{s} \in \mathcal{S}$ set

$$\mathfrak{h}(\mathfrak{s}, x) = P'(L(\mathfrak{s})|\mathcal{A}^*)(x).$$

It is easy to verify using the properties of conditional probabilities that there is a set $H \in \mathcal{A}$ with $P'(H \times 2^N) = 0$ such that whenever $x \notin H$, $\mathfrak{h}(0,x) + \mathfrak{h}(1,x) = 1$ and $\mathfrak{h}(\mathfrak{s}0,x) + \mathfrak{h}(\mathfrak{s}1,x) = \mathfrak{h}(\mathfrak{s},x)$ for $\mathfrak{s} \in \mathcal{S}$. Redefine $\mathfrak{h}(\mathfrak{s},x)$ to be $P'(X \times L(\mathfrak{s}))$ for $x \in H$ and $\mathfrak{s} \in \mathcal{S}$. Then, for each $x \in X$, the function $\mathfrak{h}(\cdot,x)$ satisfies the hypotheses of Lemma 1, so that there is a unique probability measure $T'(\cdot|x)$ on $\mathcal{B}(2^N)$ such that $T'(L(\mathfrak{s})|x) = \mathfrak{h}(\mathfrak{s},x)$ for all $\mathfrak{s} \in \mathcal{S}$. Since $\mathfrak{h}(\mathfrak{s},\cdot)$ is $\mathcal{A}$-measurable on X, so is $T'(L(\mathfrak{s})|\cdot)$ for each $\mathfrak{s} \in \mathcal{S}$, whence, by a routine measure-theoretic argument, $T'(B|\cdot)$ is $\mathcal{A}$-measurable for every $B \in \mathcal{B}(2^N)$. Denoting by $\mathcal{Q}$ the marginal of P' on X, we have for each $\mathfrak{s} \in \mathcal{S}$ and $A \in \mathcal{A}$,

$$\begin{aligned} P'(A \times L(\mathfrak{s})) &= \int_A P'(L(\mathfrak{s})|\mathcal{A}^*)(x)\, d\mathcal{Q}(x) \\ &= \int_A \mathfrak{h}(\mathfrak{s},x)\, d\mathcal{Q}(x) \quad (\text{as } \mathfrak{h}(\mathfrak{s},\cdot) = P'(L(\mathfrak{s})|\mathcal{A}^*)(\cdot) \text{ a.s. } (\mathcal{Q})) \\ &= \int_A T'(L(\mathfrak{s})|x)\, d\mathcal{Q}(x), \end{aligned}$$

so again by a routine measure-theoretic argument

$$P'(A \times B) = \int_A T'(B|x)\, d\mathcal{Q}(x)$$

for each $B \in \mathcal{B}(2^N)$. Since $P'(X \times Y) = 1$, it is easily seen that there is set $H' \in \mathcal{A}$ such that $\mathcal{Q}(H') = 0$ and $T'(Y|x) = 1$ for all $x \in X \setminus H'$. Finally, define for any $B \in \mathcal{B}(Y)$,

$$T(B|x) = \begin{cases} T'(B|x) & \text{if } x \in X \setminus H', \\ P(X \times B) & \text{if } x \in H'. \end{cases}$$

Then $T \in M(Y, \mathcal{B}(Y)|X, \mathcal{A})$ and $P = \mathcal{Q} \times T$.

Consider next the case where Y is a Borel subset of a Polish space Z. If $V_{\mathfrak{n}}$, $\mathfrak{n} \geq 1$, is a base for the toplogy of Z, we define φ: $Z \to 2^N$ by $\varphi(z) = (I_{V_1}(z), I_{V_2}(z), \ldots)$. By Lemma 2, φ is a Borel isomorphism of Z and $\varphi(Z)$, and $\varphi(Z)$ is Borel in 2^N. If we now restrict φ to Y, then (i) $\varphi(Y)$ is a Borel subset of 2^N, and (ii) φ sets up a Borel isomorphism of Y and $\varphi(Y)$. The probability measure $P \in M(X \times Y,\ \mathcal{A} \times \mathcal{B}(Y))$ can now be taken over to $(X \times \varphi(Y),\ \mathcal{A} \times \mathcal{B}(\varphi(Y)))$, factored by virtue of the previous argument, and the factorization brought back to $(X \times Y,\ \mathcal{A} \times \mathcal{B}(Y))$, again using the function φ.

Lastly, let Y be absolutely measurable and let $P \in M(X \times Y,\ \mathcal{A} \times \mathcal{B}(Y))$. Denote by μ the marginal of P on Y. As Y is absolutely measurable, there is $Y^* \subseteq Y$ such that Y^* is a Borel subset of a Polish space and $\mu(Y^*) = 1$. It follows that $P(X \times Y^*) = 1$. Restrict P to $(X \times Y^*,\ \mathcal{A} \times \mathcal{B}(Y^*))$, call it P'. By the previous paragraph, P' can be factored: $P' = \mathcal{Q} \times T'$, where $\mathcal{Q} \in M(X, \mathcal{A})$ and $T' \in M(Y^*, \mathcal{B}(Y^*)|X, \mathcal{A})$. Set $T(B|x) = T'(B \cap Y^*|x)$ for $B \in \mathcal{B}(Y)$, $x \in X$. Then $T \in M(Y, \mathcal{B}(Y)|X, \mathcal{A})$ and $P = \mathcal{Q} \times T$, completing the proof of (a)$\to$(b). (b)$\to$(c) is trivial.

To prove (c)$\to$(a), assume Y is a separable metric space, and let X be a metric completion of Y. Suppose $\lambda \in M(X)$. If $\lambda^*(Y) = 0$, Y is λ-measurable. So assume $\lambda^*(Y) > 0$, and further, by normalizing λ if necessary, assume without loss of generality that $\lambda^*(Y) = 1$. Let $\mathcal{C}$ be the smallest σ-field on X containing

$\mathcal{B}(X) \cup \{Y\}$. Extend λ to a measure μ on $\mathcal{C}$ by setting

$$\mu((E_1 \cap Y) \cup (E_2 \cap (X \setminus Y))) = \lambda(E_1), \qquad E_1, E_2 \in \mathcal{B}(X).$$

We define a probability measure P on $(X \times X, \mathcal{B}(X) \times \mathcal{C})$ by

$$P(A \times B) = \int_B I_A(\mathfrak{y})\, d\mu(\mathfrak{y}), \qquad A \in \mathcal{B}(X),\ B \in \mathcal{C},$$

so that $P(A \times B) = \mu(A \cap B)$.

Observe that the trace $\mathcal{C} \cap Y$ of the σ-field $\mathcal{C}$ on Y is just $\mathcal{B}(Y)$, so the trace of $\mathcal{B}(X) \times \mathcal{C}$ on $X \times Y$ is $\mathcal{B}(X) \times \mathcal{B}(Y)$. Furthermore, $\mu(Y) = 1$ and hence $P(X \times Y) = 1$. Let P' denote the restriction of P to $(X \times Y,\ \mathcal{B}(X) \times \mathcal{B}(Y))$, so that $P' \in M(X \times Y)$. By the hypothesis of (c), we can factor P': $P' = \mathcal{Q} \times T'$, where $\mathcal{Q} \in M(X)$ and $T' \in M(Y|X)$. We define $T(C|x) = T'(C \cap Y|x)$ for $C \in \mathcal{C}$, $x \in X$. It is now easy to verify that $T \in M(X, \mathcal{C}|X, \mathcal{B}(X))$ and that $P = \mathcal{Q} \times T$.

Next observe that for $A \in \mathcal{B}(X)$, $B \in \mathcal{C}$,

$$(*) \qquad \mu(A \cap B) = P(A \times B) = \int_A T(B|x)\, d\mathcal{Q}(x).$$

Setting $B = X$ in $(*)$, we get $\mu(A) = \mathcal{Q}(A)$, so $\mathcal{Q} = \lambda$. Setting $A = B$ in $(*)$, we have

$$\lambda(A) = \int_A T(A|x)\, d\lambda(x).$$

Since $0 \le T(A|x) \le 1$, this implies for each $A \in \mathcal{B}(X)$ that $T(A|x) = 1$ a.s. (λ) on A. So, if $\{G_{\mathfrak{n}},\ \mathfrak{n} \ge 1\}$ is a countable generating class for $\mathcal{B}(X)$, then there is a set $K \in \mathcal{B}(X)$ with $\lambda(K) = 0$ such that whenever $x \in X \setminus K$

$$T(G_{\mathfrak{n}}|x) = 1 \quad \text{if } x \in G_{\mathfrak{n}} \quad \text{and} \quad T(X \setminus G_{\mathfrak{n}}|x) = 1 \quad \text{if } x \notin G_{\mathfrak{n}}.$$

It follows immediately that for $x \notin K$, $T(\{x\}|x) = 1$. By construction of T, we have $T(Y|x) = 1$ for all $x \in X$. It follows that $X \setminus K \subset Y$, so that $\lambda_*(Y) = 1$, and hence Y is λ-measurable. This completes the proof.

References

1. M. Jirina, *Conditional probabilities on σ-algebras with countable basis*, Czechoslovak. Math. J. **4** (1954), 372–380; English transl., Selected Transl. Math. Statist. Prob., Vol. 2, Amer. Math. Soc., Providence, R.I., 1962, pp. 79–86.
2. K. Kuratowski, *Topology*, Vol. 1, 5th ed., PWN, Warsaw; Academic Press, New York; "Mir", Moscow, 1966.
3. E. Marczewski, *On compact measures*, Fund. Math. **40** (1953), 113–124.
4. Jan K. Pachl, *Disintegration and compact measures*, Math. Scand. **43** (1978), 157–168.

Department of Statistics, University of California, Berkeley, California 94720

Statistics–Mathematics Division, Indian Statistical Institute, Calcutta-700035, India