ON DECOMPOSITION INTO STOCHASTICALLY INDEPENDENT COMPONENTS

S. P. LLOYD

1. Introduction. Let $\mathcal{G}$ be an arbitrary class of measurable sets in the probability space $(\Omega, \mathcal{F}, P)$. We denote by $\mathcal{G}^*\quad$ the class of sets which are stochastically independent of $\mathcal{G}$; that is,

$$\mathcal{G}^* = \{ H \in \mathcal{F} : P(H \cap G) = P(H)P(G) \text{ for all } G \in \mathcal{G}\}.$$ 

If $\mathcal{G} \vee \mathcal{G}^*$ denotes the smallest $\sigma$-field containing $\mathcal{G}$ and $\mathcal{G}^*$, we say that decomposition (of $\mathcal{F}$ relative to $\mathcal{G}$) holds if $\mathcal{G} \vee \mathcal{G}^* = \mathcal{F}$. In other words, when decomposition holds, every statistical experiment in the class under consideration (every $\mathcal{F}$ set) can be expressed as a combination of (is in the $\sigma$-field generated by) experiments from the designated subclass (sets in $\mathcal{G}$) and experiments independent of the subclass (sets in $\mathcal{G}^*$). Examples will be given where decomposition holds and where it fails, and a sufficient condition for decomposition to hold will be established.

2. Some properties of independence. We digress briefly to exhibit some elementary properties of the $^*$ operation. We denote by $\emptyset$ the $\sigma$-field consisting of sets of probability 0 and their complements. The following relations hold for arbitrary subclasses $\mathcal{G}, \mathcal{K}, \ldots$, of $\mathcal{F}$.

(i) $\mathcal{G}^*$ is closed under the formation of complements, proper differences, and (countable) disjoint unions (but not necessarily intersections),

(ii) $\emptyset^* = \mathcal{F}$, $\mathcal{F}^* = \emptyset$,

(iii) $\mathcal{G}^* \supset \emptyset$,

(iv) If $\mathcal{G} \subset \mathcal{K}$ then $\mathcal{G}^* \supset \mathcal{K}^*$,

(v) $(\mathcal{G} \cup \mathcal{K})^* = \mathcal{G}^* \cap \mathcal{K}^*$,

(vi) $(\mathcal{G} \cap \mathcal{K})^* \subset \mathcal{G}^* \cup \mathcal{K}^*$,

(vii) $\mathcal{G} \subset \mathcal{G}^{**} = (\mathcal{G}^*)^*$,

(viii) $\mathcal{G}^{***} = \mathcal{G}^*$,

(ix) $\mathcal{G} \cap \mathcal{G}^* \subset \emptyset$,

(x) $(\mathcal{G} \cup \mathcal{G}^*)^* = \emptyset$, $(\mathcal{G} \cup \mathcal{G}^*)^{***} = \mathcal{F}$,

(xi) $(\mathcal{G} \cap \mathcal{G}^*)^* = \emptyset$, $(\mathcal{G} \cap \mathcal{G}^*)^{***} = \mathcal{F}$.

Proofs of the above properties are simple and straightforward, and are omitted. It is to be noted that the problem of when $\mathcal{G} \vee \mathcal{G}^* = \mathcal{F}$ holds is subsumed by the problem of when $\mathcal{K}^{***} = \mathcal{K}$ holds, from (xi).

Presented to the Society January 22, 1962 under the title On decomposition into independent components; received by the editors February 2, 1962.

292
3. Conditional cumulative distributions. In all that follows we assume that $\mathcal{G}$ is a sub-$\sigma$-field of $\mathcal{F}$.

**Theorem 1.** $H \in \mathcal{G}^*$ if and only if $P(H \mid \mathcal{G}) = P(H)$ with probability 1.

**Proof.** $P(H \mid \mathcal{G})$ is the unique element of $L_\infty(\Omega, \mathcal{G}, P)$ satisfying

$$P(H \cap G) = \int \! P(H \mid \mathcal{G})_\omega \, dP(\omega)$$

for all $G \in \mathcal{G}$.

When $H \in \mathcal{G}^*$ the left-hand side is equal to $P(H)P(G)$, and a solution of (2) is clearly $P(H \mid \mathcal{G}) = \text{const} = P(H)$ with probability 1. Conversely, if $P(H \mid \mathcal{G})$ is constant with probability 1 then $P(H \cap G) = P(H)P(G)$ holds for every $G \in \mathcal{G}$, whence $H \in \mathcal{G}^*$.

If $x$ is a real valued random variable we denote by $[x]$ the smallest $\sigma$-field with respect to which $x$ is Borel measurable, that is, $[x]$ is the class of sets $\{\omega: x(\omega) \in B\}$ for all linear Borel sets $B$. The conditional cumulative distribution of $x$ relative to $\mathcal{G}$ will be denoted by

$$F(x, \omega) = P(x \leq \lambda \mid \mathcal{G})_\omega, \quad -\infty < \lambda < \infty, \omega \in \Omega.$$

We can and do assume that the conditional probability on the right-hand side is determined for each $\lambda$ in such a way that $F(\cdot, \omega)$ is a bona fide cumulative distribution function for each $\omega \in \Omega$ [1, p. 29].

**Theorem 2.** Suppose $F(\lambda, \omega)$, $-\infty < \lambda < \infty$, $\omega \in \Omega$, is continuous on $-\infty < \lambda < \infty$ for each $\omega$ in the complement of a set of probability 0. Then the random variable $\xi$ defined by

$$\xi(\omega) = F(x(\omega), \omega), \quad \omega \in \Omega$$

has the properties

(i) $[\xi] \subseteq [x] \vee \mathcal{G}$,

(ii) $[x] \subseteq [\xi] \vee \mathcal{G}$,

(iii) $[\xi] \subseteq \mathcal{G}^*$.

**Remark.** P. Lévy is the first to use the independence property of random conditional cumulatives; Lévy gives no proof [2]. Since there seems to be no published proof, and since proof is not as simple as would seem off hand, we give it here.

We need first a measure theoretic lemma. Let $\mathcal{B}$ denote the $\sigma$-field of linear Borel sets, and let $(\Omega \times \Lambda, \mathcal{G} \times \mathcal{B})$ denote the product measurable space of $(\Omega, \mathcal{G})$ with $(\Lambda, \mathcal{B})$, where $\Lambda$ is the real line. If $\Gamma \subseteq \Omega \times \Lambda$ then $\Gamma_\omega$ denotes the $\omega$ section $\Gamma_\omega = \{\lambda \in \Lambda: (\omega, \lambda) \in \Gamma\}$. For any Borel set $B$, define

$$F(B, \omega) = \int_B F(d\lambda, \omega).$$
LEMMA 1. If $\Gamma \in \mathcal{G} \times \mathbb{R}$ then

$$P(x(\omega) \in \Gamma_\omega) = \int F(\Gamma_\omega, \omega)dP(\omega), \quad \Gamma \in \mathcal{G} \times \mathbb{R}. \tag{3}$$

PROOF. The properties of $F(\cdot, \cdot)$ insure that $\Delta_1(\Gamma) = \int F(\Gamma_\omega, \omega)dP(\omega)$ is a measure on $\mathcal{G} \times \mathbb{R}$. With $\delta: \Omega \rightarrow \mathcal{G} \times \mathbb{R}$ defined by $\delta(\omega) = (\omega, x(\omega))$, let $\Delta_2$ be the measure $\Delta_2(\Gamma) = P(\delta^{-1}(\Gamma))$, $\delta^{-1}(\Gamma) \in \mathcal{F}$, noting that $\Delta_2(\Gamma) = P(x(\omega) \in \Gamma_\omega)$. When $\Gamma$ is a $\mathcal{G} \times \mathbb{R}$ rectangle, $\Gamma = G \times B$, $G \in \mathcal{G}$, $B \in \mathbb{R}$, we have

$$\Delta_2(\Gamma) = P(G \cap \{ x \in B \})$$
$$= \int_B P(x \in B \mid G) \omega dP(\omega)$$
$$= \int F(\Gamma_\omega, \omega)dP(\omega) = \Delta_1(\Gamma).$$

Since measures $\Delta_1$ and $\Delta_2$ coincide on $\mathcal{G} \times \mathbb{R}$ rectangles, they coincide on $\mathcal{G} \times \mathbb{R}$, proving the lemma.

PROOF OF THEOREM 2. Define the pseudoinverses of $F(\cdot, \omega)$ as

$$\overline{G}(u, \omega) = \sup \lambda: F(\lambda, \omega) \leq u$$
$$\underline{G}(u, \omega) = \min \lambda: F(\lambda, \omega) \geq u,$$

$$0 < u < 1, \omega \in \Omega.$$

We have, for all values of the arguments,

$$F(\lambda, \omega) < u \quad \text{if and only if} \quad \lambda < \overline{G}(u, \omega), \quad 0 < u < 1, \quad -\infty < \lambda < \infty, \omega \in \Omega. \tag{4}$$

The assumed continuity of $F(\cdot, \omega)$ gives

$$F(\overline{G}(u, \omega), \omega) = F(\underline{G}(u, \omega), \omega) = u, \quad 0 < u < 1,$$

with probability 1.

We will show that the conditional distribution of $\xi$ given $\mathcal{G}$ is the uniform-on-the-unit-interval distribution with probability 1. For each $G \in \mathcal{G}$ and each $u$, $0 < u < 1$, we have

$$P(G \cap \{ \xi < u \}) = P(G \cap \{ F(x(\omega), \omega) < u \})$$
$$= P(G \cap \{ x(\omega) < \overline{G}(u, \omega) \})$$
$$= P(x(\omega) \in \Gamma_\omega),$$

where $\Gamma \subseteq \Omega \times \mathbb{R}$ is the set.
\[ \Gamma = (G \times \Lambda) \cap \Delta, \]
\[ \Delta = \{ (\omega, \lambda) : \lambda < G(u, \omega) \}. \]

It follows from (4) that \( G(u, \cdot) \) is \( \mathcal{G} \) measurable for each fixed \( u, \) \( 0 < u < 1; \) to see that \( \Delta \) is \( \mathcal{G} \times \mathcal{A} \) measurable, observe that \( \Delta = \lim_n \Delta_n, \)

\[ \Delta_n = \bigcup_{j=-\infty}^{\infty} D_{n,j} \times (-\infty, j/2^n) \]
\[ D_{n,j} = \{ \omega : j/2^n < G(u, \omega) \leq (j + 1)/2^n \}, \]

and the approximating sequence is manifestly \( \mathcal{G} \times \mathcal{A} \) measurable. Applying Lemma 1, we have, for each \( G \in \mathcal{G} \) and each \( u, 0 < u < 1, \)

\[ \int_{\mathcal{G}} P(\xi < u | \mathcal{G}) dP(\omega) = P(G \cap \{ \xi < u \}) = P(x(\omega) \in \Gamma_u) \]
\[ = \int_{\mathcal{G}} F(G(\omega, \omega) - 0, \omega) dP(\omega) \]
\[ = uP(G), \quad G \in \mathcal{G}, \quad 0 < u < 1, \]

whence \( P(\xi < u | \mathcal{G}) = u \) with probability 1, and this for each \( u, \) \( 0 < u < 1. \)

That \( \xi \) is \( [x] \cap \mathcal{G} \) measurable follows easily from (4). It remains to show that \( x \) is \( [\xi] \cap \mathcal{G} \) measurable. Define random variables \( x \) and \( \hat{x} \) by

\[ x(\omega) = G(\xi(\omega), \omega) \]
\[ \hat{x}(\omega) = G(\xi(\omega), \omega), \]

(where \( \omega \in \Omega \)), observing that \( \hat{x} \) is \( [\xi] \cap \mathcal{G} \) measurable, from (4), and that \( x(\omega) \leq \hat{x}(\omega) \) holds for each \( \omega \in \Omega. \) We will show that \( x = \hat{x} \) with probability 1. For each \( \omega \in \Omega \) let \( I(\omega) \) denote the union of the disjoint nontrivial closed intervals on which \( F(\cdot, \omega) \) is constant. Since \( \{ \omega : \hat{x}(\omega) < x(\omega) \} \subset \{ \omega : x(\omega) \in I(\omega) \}, \) it will suffice to prove that \( P(x(\omega) \in I(\omega)) = 0. \) If Lemma 1 is applicable we have

\[ P(x(\omega) \in I(\omega)) = \int F(I(\omega), \omega) dP(\omega) = 0, \]

under the assumption that the \( \Omega \times \Lambda \) set \( \Gamma \) whose \( \omega \) sections are \( \Gamma_{\omega} = I(\omega) \) is \( \mathcal{G} \times \mathcal{A} \) measurable. To see that this is the case, observe that \( \Gamma = \lim_n \Gamma_n, \) with

\[ \Gamma_n = \bigcup_{-\infty < j < k < \infty} G_{n,j,k} \times [(j - 1)/2^n, (k + 1)/2^n], \]
\[ G_{n,j,k} = \{ \omega : F((j - 1)/2^n, \omega) < F(j/2^n, \omega) \}
\[ = F(k/2^n, \omega) < F((k + 1)/2^n, \omega) \}.


Our main result is the following, to the effect that if there exists one such random variable $\xi$ then there exist many.

**Theorem 3.** Suppose that there exists a random variable independent of $\mathcal{G}$ whose distribution is continuous. Then decomposition holds; that is, $\mathcal{G} \cap \mathcal{G}' = \mathcal{F}$.

**Proof.** From Theorem 2, we may assume given a random variable $\xi$ with the property $P(\xi < u | \mathcal{G}) = u$ with probability 1, $0 < u < 1$. For arbitrary $A \in \mathcal{F}$ consider the random variable $\eta_A = \chi_A + \xi$, where $\chi_A$ denotes the characteristic function of set $A$. The conditional distribution of $\eta_A$ has the property

$$P(\lambda_1 \leq \eta_A < \lambda_2 | \mathcal{G}) = P(A \cap [\lambda_1 - 1 \leq \xi < \lambda_2 - 1] | \mathcal{G})$$

$$+ P(A' \cap [\lambda_1 \leq \xi < \lambda_2] | \mathcal{G}) \leq 2(\lambda_2 - \lambda_1)$$

with probability 1, implying that some version of $P(\eta_A < \lambda | \mathcal{G})$ is continuous in $\lambda$ with probability 1. Applying Theorem 2 to $\eta_A$, we obtain a random variable $\xi_A$ such that $\eta_A$ is $[\xi_A] \cap \mathcal{G}$ measurable and such that $[\xi_A] \subset \mathcal{G}'$. Since $0 < \xi < 1$ with probability 1, set $A$ differs from the set $\{\eta_A > 1\}$ by a set of probability 0, whence $A \subset [\eta_A \cap \mathcal{G}] \cap [\xi_A \cap \mathcal{G}' \cap \mathcal{G}]$ modulo set of probability 0. But $A$ is an arbitrary element of $\mathcal{F}$, so that $\mathcal{F} \subset \mathcal{G} \cap \mathcal{G}'$, or since the reverse inclusion is trivial, $\mathcal{F} = \mathcal{G} \cap \mathcal{G}'$.

4. Examples.

**Example 1.** Decomposition holds. Let $(\Omega, \mathcal{F}, P)$ be atomless, and let $\mathcal{G}$ be purely atomic. That is, there is a finite or countable partition of $\Omega$ into disjoint sets $G_i$, $P(G_i) > 0$, $i = 1, 2, \ldots$, and $\mathcal{G}$ is the $\sigma$-field generated by the $G_i$, $i = 1, 2, \ldots$. From Theorem 1, $H \in \mathcal{G}'$ if and only if $P(H \cap G_i) / P(G_i)$ is independent of $i$. To see that decomposition holds, let $A \in \mathcal{F}$ be an arbitrary measurable set, and define $H_{ii} = A \cap G_i$, $i = 1, 2, \ldots$. Since $(\Omega, \mathcal{F}, P)$ is atomless, for each $i$ there exist for all $j$ sets $H_{ij} \subset G_j$ such that $P(H_{ij}) / P(G_j) = P(H_{ij}) / P(G_i)$. With $H_i = \bigcup_j H_{ij}$, we have $H_i \in \mathcal{G}'$, $i = 1, 2, \ldots$, and finally $A = \bigcup_i H_{ii} = \bigcup_i (G_i \cap H_i) \subset \mathcal{G} \cap \mathcal{G}'$. We will prove also that $\mathcal{G} = \mathcal{G}''$ modulo sets of probability 0. Suppose $A \in \mathcal{F}$ does not differ from some $\mathcal{G}$ set by a set of probability 0. Then both $P(A \cap H_i) > 0$ and $P(A' \cap G_i) > 0$ hold for at least one $i$, say $i = j$. Choose $H_j \subset A \cap G_j$ and $K_j \subset A' \cap G_j$ so that $P(H_j) = P(K_j) > 0$. For all $k \neq j$ choose sets $H_k = K_k \subset G_k$ so that $P(H_k) / P(G_k) = P(H_j) / P(G_j)$. Defining $H = \bigcup_i H_i$, $K = \bigcup_i K_i$, we have $H \subset \mathcal{G}'$ and $P(H) = P(K)$. Since $P(A \cap H) - P(A \cap K) = 2P(H_j) > 0$, at least one of $H, K$ is not independent of $A$, implying $A \not\in \mathcal{G}''$. Thus $\mathcal{G}'' \subset \mathcal{G}$ modulo sets of probability 0, whence $\mathcal{G}'' = \mathcal{G}$ modulo sets of probability 0, from (1.vii).
Example 2. Decomposition fails. Let \( \Omega = \{(x, i) : 0 \leq x \leq 1, i = 1, 2\} \), and let probability have density \( p_i(x) \) at \((x, i)\) relative to linear Borel measure on the two intervals which comprise \( \Omega \). Assume that \( p_1(x) + p_2(x) = 1, 0 \leq x \leq 1 \). If \( B_1, B_2 \) are Borel subsets of the unit interval we denote by \( [B_1, B_2] \) the \( \omega \) set \( \{(x, 1) : x \in B_1\} \cup \{(x, 2) : x \in B_2\} \). Let \( \mathcal{G} \) be the sub-\( \sigma \)-field of sets of the form \([B, B]\) for all Borel \( B \). It is readily verified that conditional probabilities relative to \( \mathcal{G} \) have the form

\[
P([B_1, B_2] \mid \mathcal{G})_{(x,i)} = x_{B_1}(x)p_1(x) + x_{B_2}(x)p_2(x),
\]

\(0 \leq x \leq 1, i = 1, 2\).

For each \( \omega \) the measure \( P(\cdot \mid \mathcal{G})_\omega \) is purely atomic, the measures of the two atoms being \( p_1(x) \), \( p_2(x) \). It is easy to choose \( p_1 \), \( p_2 \) in such a way that the only constant functions of the form (5) are 0 and 1, which is to say, \( \mathcal{G}^* = \emptyset \). Thus decomposition fails. Note also that \( \mathcal{G}^* = \emptyset \neq \mathcal{G} \).

Example 2 suggests the following conjecture: for decomposition to hold it is sufficient that the conditional probability \( P(\cdot \mid \mathcal{G}) \) be an atomless measure with probability 1. (This statement makes sense in the Stone representation space; see [3].)

5. Acknowledgment. Example 2 was suggested by Professor Murray Rosenblatt, and the author wishes to express his thanks to Professor Rosenblatt for a number of interesting conversations on this problem.

References


Bell Telephone Laboratories,
Murray Hill, New Jersey