ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 3, Number 4, Fall 1973

EXCHANGEABLE EVENTS AND COMPLETELY MONOTONIC SEQUENCES

CLARK H. KIMBERLING [1]

1. **Introduction.** We investigate the probability theory of an infinite sequence of events, all having the same probability, and we assume the events are exchangeable. That is, for each positive integer n, each pair of n-fold intersections of events have the same probability. In the familiar game of tossing a coin forever, the events "heads" provide an example of a sequence of exchangeable events. A less trivial example is provided by Pólya's urn model. (See, for example, Feller [**1**, p. 226].)

We show in Theorem 1 that a sequence of events of common probability is exchangeable if and only if the sequence of real numbers whose nth term is the probability common to the n-fold intersections of events is a completely monotonic sequence. Theorem 2 asserts that for such events, Kolmogorov's Strong Law of Large Numbers holds if and only if the events are independent. Theorems 3, 4, 5, and 6 describe probabilities of unions and intersections of exchangeable events of common probability.

We recall now preliminaries from Feller [**1**, p. 225], Widder [**4**, p. 108 and p. 12], and Hardy [**2**, pp. 279–282]. For a sequence $\mu_0, \mu_1, \cdots$ of real numbers, we denote by $\Delta^m \mu_q$ the sum

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} \mu_{q+m-k}$$

and define $\mu_0, \mu_1, \cdots$ to be completely monotonic if

$$(-1)^m \Delta^m \mu_q \geqq 0 \quad \text{for } m, q = 0, 1, \cdots,$$

and minimally completely monotonic if for every $\mu < \mu_0$ the sequence $\mu, \mu_1, \mu_2, \cdots$ is not completely monotonic.

A fundamental theorem of Hausdorff is that $\mu_0, \mu_1, \cdots$ is completely monotonic if and only if there exists a bounded nondecreasing function α from $[0, 1]$ into $[0, \infty)$ such that

Received by the editors June 28, 1971 and, in revised form, March 20, 1972.
AMS (MOS) subject classifications (1970). Primary 39A05, 60G05, 40—01; Secondary 26A48, 60—01.
[1]The author is grateful for support from the University of Evansville and from NSF Grant GY5595.

$$\mu_n = \int_0^1 t^n \, d\alpha(t) \quad \text{for } n = 0, 1, \cdots.$$

We always take α to be the unique normalized function of Widder [4, p. 100] and refer to α as the distributor for $\mu_0, \mu_1, \cdots$.

The essentials of the proof of the next theorem, referred to in the sequel as Hardy's theorem, may be found in Hardy [2] and Widder [4]. Suppose $\mu_0, \mu_1, \cdots$ is a completely monotonic sequence with distributor α. Then the following are equivalent:

(i) There exists μ such that the sequence $\mu, \mu_0, \mu_1, \cdots$ is completely monotonic.

(ii) $\sum_{i=0}^{\infty} (-1)^i \Delta^i \mu_0 < \infty$.

(iii) There exists a distributor γ such that

$$\alpha(t) = \int_0^t s \, d\gamma(s);$$

in this case, γ is the distributor of $\mu, \mu_0, \mu_1, \cdots$.

(iv) $\int_0^1 (1/t) \, d\alpha(t) < \infty$.

That μ for which $\mu, \mu_0, \mu_1, \cdots$ is minimally completely monotonic, when such a μ exists, is

$$\sum_{i=0}^{\infty} (-1)^i \Delta^i \mu_0 = \int_0^1 \frac{1}{t} \, d\alpha(t).$$

2. **Main results.** We define a sequence of real numbers $\mu_0, \mu_1, \mu_2, \cdots$, where $\mu_0 = 1$, to be admissible if there exists a probability space $(\Omega, \mathcal{A}, P)$ and events A_n in $\mathcal{A}$ such that

$$P(A_{n_1} \cdots A_{n_m}) = \mu_m$$

whenever

$$1 \leqq n_1 < \cdots < n_m \quad \text{for } m = 1, 2, \cdots.$$

We say that the events A_n and their indicators I_n are exchangeable with respect to $\mu_0, \mu_1, \cdots$.

THEOREM 1. *Let $\mu_0, \mu_1, \cdots$ be a sequence with $\mu_0 = 1$. Then the sequence $\mu_0, \mu_1, \cdots$ is admissible if and only if it is completely monotonic.*

PROOF. Granted admissibility, complete monotonicity is immediate by de Finetti's theorem (see, for example, Feller [1, p. 225]).

To prove the converse, we define for $m = 1, 2, \cdots$ an m-place function $F_{1,\cdots,m}$ by

$$F_{1,\cdots,m}(x_1, \cdots, x_m) = \begin{cases} 0 & \text{if } \min\{x_1, \cdots, x_m\} < 0, \\ \mu_{m-j} & \text{if } \min\{x_1, \cdots, x_m\} \geqq 0, \\ & \text{if exactly } m-j \text{ of these coordinates are in } [0,1) \text{ and} \\ & j \text{ coordinates are} \geqq 1. \end{cases}$$

For each finite collection $n_1 < \cdots < n_k$ of indices, define $F_{n_1,\cdots,n_k}$ marginally, viz.,

$$F_{n_1,\cdots,n_k}(x_1, \cdots, x_k) = F_{1,2,\cdots,n_k}(1, \cdots, 1, x_1, 1, \cdots, \cdots, 1, x_k),$$

where x_j occupies the n_jth place for $j = 1, \cdots, k$. For each permutation $\lambda_{n_1}, \cdots, \lambda_{n_k}$ of $n_1, \cdots, n_k$, define

$$F_{\lambda_{n_1},\cdots,\lambda_{n_k}} = F_{n_1,\cdots,n_k}.$$

Then

$$F_{\lambda_1,\cdots,\lambda_k} = F_{1,\cdots,k}$$

for every set of k distinct indices $\lambda_1, \cdots, \lambda_k$.

The collection

$$\mathfrak{F} = \{F_{\lambda_1,\cdots,\lambda_n} : \{\lambda_1, \cdots, \lambda_n\} \text{ is a finite collection of positive integers}\}$$

clearly satisfies items a, b, c, e, and f of the hypothesis of the Kolmogorov theorem, as presented in Tucker [3, p. 30]. We proceed with an inductive argument to show that item d is also satisfied. Letting $\mu_{F_1,\cdots,m}$ be the probability measure induced on R^m by $F_{1,\cdots,m}$, we wish to show that for every m dimensional cell $(a, b] = ((a_1, \cdots, a_m), (b_1, \cdots, b_m)]$, we have

$$\mu_{F_1,\cdots,m}(a, b] = \sum_{k=0}^{m} (-1)^k \sum_{\delta \in \Delta_{k,m}} F_{1,\cdots,m}(\delta) \geqq 0, \tag{1}$$

where δ ranges through the set $\Delta_{k,m}$ of $\binom{m}{k}$ vertices of $(a, b]$ which consist of k a_i's and $m - k$ b_i's, for $m = 1, 2, \cdots$.

For $m = 1$, (1) is just $F_1(b) - F_1(a) \geqq 0$. Now fix $m > 1$, and suppose that (1) holds in $\mathfrak{F}$ for all cells $(a, b]$ of dimension $m - 1$. For $q = m$, we shall write just F for $F_{1,\cdots,q}$. Let

$$(a, b] = ((a_1, \cdots, a_m), (b_1, \cdots, b_m)]$$

be an arbitrary m dimensional cell.

Lemma. *Suppose that for some j satisfying $1 \leqq j \leqq m$, one of the following is true:*

$$a_j \leqq b_j < 0, \quad 0 \leqq a_j \leqq b_j < 1, \quad or \quad 1 \leqq a_j \leqq b_j.$$

Then we see that (1) *holds with equality, since the terms in the sum in* (1) *can be paired as*

$$F(x_1, \cdots, b_j, \cdots, x_m) - F(x_1, \cdots, a_j, \cdots, x_m),$$

and each such difference is equal to zero.

Now let q be the number of components b_i of b such that $b_i < 1$. If the hypothesis of the lemma does not hold, we have $a_i < 0$ for those i satisfying $b_i < 1$.

Case 1. If the remaining $m - q$ a_i's satisfy $0 \leqq a_i < 1$, then

$$F(\delta) = \begin{cases} \mu_q & \text{if } \delta = b, \\ \vdots & \\ \mu_{q+k} & \text{if } \delta \text{ has exactly } k \text{ nonnegative } a_i\text{'s}, \\ \vdots & \\ \mu_m & \text{if } \delta \text{ has } m - q \text{ nonnegative } a_i\text{'s}, \\ 0 & \text{otherwise, since in all the remaining cases, at least one component of } \delta \text{ is negative.} \end{cases}$$

If $0 \leqq k \leqq m - q$, then $\binom{m-q}{k}$ is the number of vertices having exactly k nonnegative a_i's, so that

$$\sum_{\delta \in \Delta_{k,m}} F(\delta) = \binom{m-q}{k} \mu_{q+k}.$$

Thus,

$$\begin{aligned} \sum_{k=0}^{m} (-1)^k \sum_{\delta' \in \Delta_{k,m}} F_{1,\cdots,m}(\delta) &= \sum_{k=0}^{m-q} (-1)^k \sum_{\delta \in \Delta_{k,m}} F_{1,\cdots,m}(\delta) \\ &= \sum_{k=0}^{m-q} (-1)^k \binom{m-q}{k} \mu_{q+k} \\ &= (-1)^{m-q} \sum_{k=0}^{m-q} (-1)^k \binom{m-q}{k} \mu_{m-k} \\ &= (-1)^{m-q} \Delta^{m-q} \mu_q \geqq 0. \end{aligned}$$

Case 2. If the lemma does not apply and if for some k satisfying $1 \leqq k \leqq m$, we have $b_k \geqq 1$ and $a_k < 0$, then we define

$$a' = (a_1, \cdots, a_{k-1}, a_{k+1}, \cdots, a_m)$$

and

$$b' = (b_1, \cdots, b_{k-1}, b_{k+1}, \cdots, b_m)$$

and claim that

$$\mu_F(a, b] = \mu_{F_{1,\cdots,m-1}}(a', b'],$$

this second expression being nonnegative by the induction hypothesis. To see that

$$\mu_F(a, b] = \mu_{F_{1,\cdots,m-1}}(a', b'],$$

define for each vertex δ of $(a, b]$ the vertex δ' of $(a', b']$ formed by removal of the kth component of δ. Then

$$F(\delta) = \begin{cases} F_{1,\cdots,m-1}(\delta') & \text{if the } k\text{th component of } \delta \text{ is } b_k, \\ 0 & \text{if the } k\text{th component of } \delta \text{ is } a_k. \end{cases}$$

Now decomposing $\Delta_{k,m}$ into

$$\Delta(b) = \{\delta \in \Delta_{k,m} : b_k \text{ is the } k\text{th component of } \delta\}$$

and

$$\Delta(a) = \{\delta \in \Delta_{k,m} : a_k \text{ is the } k\text{th component of } \delta\},$$

we have

$$\sum_{\delta \in \Delta_{k,m}} F(\delta) = \sum_{\delta \in \Delta(b)} F(\delta) + \sum_{\delta \in \Delta(a)} F(\delta) = \sum_{\delta' \in \Delta_{k,m-1}} F_{1,\cdots,m-1}(\delta'),$$

so that

$$\sum_{k=1}^{m} (-1)^k \sum_{\delta \in \Delta_{k,m}} F(\delta) = \sum_{k=1}^{m-1} (-1)^k \sum_{\delta \in \Delta_{k,m-1}} F_{1,\cdots,m-1}(\delta')$$
$$= \mu_{F_{1,\cdots,m-1}}(a', b'].$$

This completes a proof that item d of the hypothesis of the Kolmogorov theorem is satisfied by the collection $\mathfrak{F}$. Applying the Kolmogorov theorem, we now conclude that there exist a probability space $(\Omega, \mathcal{A}, P)$ and random variables ξ_n over Ω whose distribution functions and joint distribution functions are, with corresponding indices, those in $\mathfrak{F}$. In particular

$$F_{\xi_n}(x) = \begin{cases} 0 & \text{if } x < 0, \\ \mu_1 & \text{if } 0 \leqq x < 1, \\ 1 & \text{if } x \geqq 1 \end{cases}$$

and letting $A_n = \xi_n^{-1}(-\infty, 0]$, we have for every positive integer m, and indices $n_1 < \cdots < n_m$,

$$P\left(\bigcap_{j=1}^{m} A_{n_j}\right) = P\left(\bigcap_{j=1}^{m} \xi_{n_j}^{-1}(-\infty, 0]\right)$$
$$= F_{n_1, \cdots, n_m}(0, \cdots, 0) = F_{1, \cdots, m}(0, \cdots, 0) = \mu_m.$$

Therefore, the sequence $\mu_0, \mu_1, \cdots$ is admissible.

THEOREM 2. *Let $\mu_0, \mu_1, \cdots$ be a sequence admissible with respect to indicators $I_1, I_2, \cdots$. Then there exists a constant random variable c such that the sequence*

$$I_1, \tfrac{1}{2}(I_1 + I_2), \cdots, (I_1 + \cdots + I_n)/n, \cdots$$

of arithmetic means converges in probability to c if and only if the indicators $I_1, I_2, \cdots$ are independent, in which case the random variable c is given by

$$c(\omega) = \mu_1 \quad \textit{for all } \omega \in \Omega.$$

PROOF. It is well known (for example, Tucker [3, pp. 123–124]) that if the I_n are independent, then the arithmetic means converge not only in probability to μ_1, but, *a fortiori*, with probability one.

To prove the converse, let α_n denote the distribution function of the random variable $\zeta_n = (I_1 + \cdots + I_n)/n$ and assume that for some constant random variable c, we have

$$\lim_{n\to\infty} P[\,|\zeta_n - c| \geqq \epsilon] = 0 \quad \text{for every } \epsilon > 0.$$

Then

$$\lim_{n\to\infty} P[c - \epsilon \leqq \zeta_n \leqq c + \epsilon] = 1,$$

so that

$$\lim_{n\to\infty} P[\zeta_n \leqq t] = \begin{cases} 0 & \text{if } t < c, \\ 1 & \text{if } t \geqq c. \end{cases}$$

This shows that

$$\lim_{n\to\infty} \alpha_n(t) = \begin{cases} 0 & \text{if } t < c, \\ 1 & \text{if } t \geqq c. \end{cases} \tag{2}$$

Now let α be the distributor of the sequence $\mu_0, \mu_1, \cdots$. As proved in Feller [1, p. 223],

$$\lim_{n\to\infty} \alpha_n(t) = \alpha(t)$$

at each point t of continuity of α. As a nondecreasing function, α has at most countably many points of discontinuity. Letting ρ be an arbitrarily small positive number, we can therefore find points p_1 and p_2 of continuity of α such that $c - \rho < p_1 < c < p_2 < c + \rho$. It follows from (2) and our normalization agreement for distributors that

$$\alpha(t) = \begin{cases} 0 & \text{if } t < c, \\ \frac{1}{2} & \text{if } t = c, \\ 1 & \text{if } t > c. \end{cases}$$

Thus, from the representation

$$\mu_n = \int_0^1 t^n \, d\alpha(t),$$

we obtain $\mu_n = c^n$. But this means that if m is any positive integer and $n_1 < \cdots < n_m$ any m indices, then

$$P(A_{n_1} \cdots A_{n_m}) = P(A_{n_1}) \cdots P(A_{n_m}).$$

Therefore, the sets $A_1, A_2, \cdots$, and consequently, the corresponding indicators $I_1, I_2, \cdots$, are independent.

THEOREM 3. *Suppose $\mu_0, \mu_1, \cdots$ is admissible with respect to a probability space $(\Omega, \mathcal{A}, P)$, sets A_n in $\mathcal{A}$, and distributor α. Then the A_n satisfy the converse of the Borel-Cantelli Lemma (as in Tucker [3, p. 70]) if and only if $\alpha(0+) = 0$. In fact, $P[A_n \text{ i.o.}] = 1 - \alpha(0+)$.*

PROOF. Since

$$P\left(\bigcap_{j=m}^{n} A_j^c \right) = \int_0^1 (1-t)^{n-m+1} \, d\alpha(t)$$

we obtain

$$\lim_{n\to\infty} P\left(\bigcap_{j=m}^{n} A_j^c\right) = \alpha(0+).$$

Now,

$$P[A_n \text{ i.o.}] = P\left(\bigcap_{m=1}^{\infty}\bigcup_{j=m}^{\infty} A_j\right)$$

$$= \lim_{m\to\infty} P\left(\bigcup_{j=m}^{\infty} A_j\right) = 1 - \lim_{m\to\infty} P\left(\bigcap_{j=m}^{\infty} A_j^c\right)$$

$$= 1 - \lim_{m\to\infty}\lim_{n\to\infty} P\left(\bigcap_{j=m}^{n} A_j^c\right)$$

$$= 1 - \lim_{m\to\infty} \alpha(0+) = 1 - \alpha(0+).$$

THEOREM 4. *Suppose* $\mu_0, \mu_1, \cdots$ *is admissible with respect to sets* $A_1, A_2, \cdots$ *and distributor* α. *Then*

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] = P[A_n \textit{ i.o.}].$$

Thus, with probability one, if a point lies in any A_n, *then it lies in infinitely many* A_n*'s.*

PROOF.

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] = P(A_1) + P(A_2 - A_1) + P(A_3 - (A_2 \cup A_1)) + \cdots$$

$$= P(A_1) + P(A_2A_1^c) + P(A_3A_2^cA_1^c) + \cdots$$

$$= \sum_{n=1}^{\infty} (-1)^{n-1}\Delta^{n-1}\mu_1$$

by de Finetti's theorem

$$= 1 - \alpha(0+)$$

by Hardy's theorem

$$= P[A_n \text{ i.o.}]$$

by Theorem 3.

THEOREM 5. *If* $P(\bigcup_{n=1}^{\infty} A_n) = 1$ *and* $P(\bigcap_{n=1}^{\infty} A_n) = 0$, *or, equivalently, if* $\alpha(0+) = 0$ *and* $\alpha(1-) = 1$, *then with probability one, a point* *of Ω lies in infinitely many A_n's and in infinitely many A_n^c's.*

PROOF. If

$$P\left(\bigcap_{n=1}^{\infty} A_n\right)= 0,$$

then

$$P\left(\bigcup_{n=1}^{\infty} A_n^c\right)= 1.$$

By Theorem 4

$$P[A_n^c \text{ i.o.}] = 1.$$

Also by Theorem 4

$$P[A_n \text{ i.o}] = 1,$$

since

$$P\left(\bigcup_{n=1}^{\infty} A_n\right)= 1.$$

THEOREM 6. *Suppose $\mu_0, \mu_1, \cdots$ is a minimally completely monotonic sequence with $\mu_1 < 1$, with respect to sets $A_1, A_2, \cdots$. Then for every finite collection $A_{i_1}, \cdots, A_{i_n}$, $P(\bigcup_{j=1}^{n} A_{i_j}) < 1$, while for every infinite collection $A_{i_1}, A_{i_2}, \cdots, P(\bigcup_{j=1}^{\infty} A_{i_j}) = 1$. That is, no finite collection of A_i's covers Ω, while every infinite collection does cover Ω, with probability one.*

PROOF.

$$P\left[\bigcup_{j=1}^{n} A_{i_j}\right] = \sum_{i=0}^{n-1} (-1)^i \Delta^i \mu_1$$

$$< \sum_{i=0}^{\infty} (-1)^i \Delta^i \mu_1 = \mu_0 = 1,$$

while

$$P\left(\bigcup_{j=1}^{\infty} A_{i_j}\right)= \sum_{i=0}^{\infty} (-1)^i \Delta^i \mu_1 = 1$$

by Hardy's theorem.

References

1. W. Feller, *An introduction to probability theory and its applications.* Vol. II, Wiley, New York, 1966. MR **35** #1048.

2. G. H. Hardy, *Divergent series,* Clarendon Press, Oxford, 1949. MR **11**, 25.

3. H. G. Tucker, *A graduate course in probability,* Probability and Math Statist., vol. 2, Academic Press, New York, 1967. MR **36** #4593.

4. D. V. Widder, *The Laplace transform,* Princeton Math. Series, vol. 6 Princeton Univ. Press, Princeton, N. J., 1946. MR **3**, 232.

University of Evansville, Evansville, Indiana 47701