THE PHILIPPINE STATISTICIAN
Volume XVI, Nos. 3/4, 1967

# MATRIX-VALUED PROBABILITY THEORY[1]

by
F. M. SIOSON

1. **Introduction.** The paper deals with the mathematical foundations of a probability theory not hitherto considered in the literature. It follows the axiomatic approach proposed by A. N. Kolmogorov [2] in (real-valued) ordinary probability theory. While this point of view may not be as intuitively sound and logically satisfactory as those proposed later by Jerzy Los [3] and Yukiyosi Kawada ]1[, it nevertheless is the most well-known if not the most elementary and easily developed.

Like all mathematical theories, probability theory may be founded on the theory of sets, which will consequently be assumed here. Kolmogorov's approach starts with a set U (called the population or sample space). Intuitively, the elements of U consist of all the possible outcomes of a random experiment under consideration. A (random) event is a subset of U, but, for reasons of both a practical and theoretical nature, not every subset of U is, in general, an event. The fundamental requirement, in any case, for any family of subsets of U to be an admissible family of events over U is that it forms a Boolean algebra under the set-theoretical operations of union $\cup$, inter-section $\cap$, and complementation $'$. For our purposes, the following is a sufficient requirement.

DEFINITION 1. A family F of subsets of U is a Boolean algebra (that is, a field of events over U) if and only if

(a) $\phi, U \in F$,

[1] This communication is an excerpt from a body of results obtained by author while working as a consultant to the Bureau of the Census and Statistics, Manila. The author is a Professor and Chairman of the Mathematics Department at the Ateneo.

(b) if $X, Y \in F$, then $X - Y \in F$ and $X \cap Y \in F$.

The same family F is called a $\sigma$ – algebra (or a $\sigma$ – field ot events over U) if and only if in addition

(c) $X_i \in F$ $(i = 1, 2, \ldots, n, \ldots)$ implies $\bigcap_{i=1}^{\infty} X_i \in F$.

Clearly, if F is a field of events ($\sigma$ – field of events) over U, then

(b'.) $X, Y \in F$ implies $X \cup Y = U - [(U - X) \cap (U - Y)] \in F$, (c') $X_i \in F$ for $i = 1, 2, \ldots, n, \ldots$ implies $\bigcup_{i=1}^{\infty} X_i = U - [\bigcap_{i=1}^{\infty} (U - X_i)] \in F$).

From matrix theory, recall that an n by n real symmetric matrix A is said to be positive semi-definite if and only of for every real row vector x we have $xAx' \geqq 0$. Let us denote $A \geqq 0$ when A is positive semi-definite and $A \geqq B$ if and only $A - B \geqq 0$. Denote by $[0, I]$ the set of all positive semi-definite n by n matrices A such that $0 \leqq A \leqq I$, where I is the n by n identity matrix.

DEFINITION 2. A matrix-valud probability space is a triple (U, F, P) consisting of a sample space U, a $\sigma$ — field F of events over U, and a set function $P: F \longrightarrow [O, I]$ such that

(a) $P(U) = I$,

(b) P is countably additive, i.e. for any family of pairwise disjoint subsets $X_i \in F$ $(i = 1, 2, \ldots, n, \ldots)$, $P(\bigcup_{i=1}^{\infty} X_i) = \sum_{i=1}^{\infty} P(X_i)$.

2. **The Fundamental Theorem.** We will need a couple of Lemmata to prove the fundamental theorem.

LEMMA A. If F is a $\sigma$ — field of events over U and $P_{ij}$: $F \longrightarrow R$ is a real-valued bounded countably additive set function, then there exists an element $M \epsilon F$ such that $P_{ij}(M)$ is maximum (similarly for minimum).

Proof. We shall only prove the former result. The proof of the parenthetical remark follows in a similar manner. Let $m = \sup [P_{ij}(X) : X \epsilon F]$.

First note that $P_{ij}(\phi) = P_{ij}(\phi \cup \phi \cdot \cup \ldots \cup \phi \cup \ldots) = P_{ij}(\phi) + P_{ij}(\phi \cup \ldots \cup \phi \cup \ldots) = P_{ij}(\phi) + P_{ij}(\phi) = 2P_{ij}(\phi)$ and hence $P_{ij}(\phi) = O$.

Thus, clearly the number m is non-negative. Let $X_1, X_2, \ldots, X_n, \ldots$ be elements belonging to F such that $\lim_{n\to\infty} P_{ij}(X_n) = m$. Since $P_{ij}$ is bounded, then all $P_{ij}(X_n)$ are finite. Also observe that

$$X_n = \bigcup_{k=n}^{\infty} (X_k - X_{k+1}) \cup \bigcap_{k=1}^{\infty} X_k,$$

where the sets occurring on the right are also pairwise disjoint. Thus

$$P_{ij}(X_n) = \sum_{k=n}^{\infty} P_{ij}(X_k - X_{k+1}) + P_{ij}\left(\bigcap_{k=1}^{\infty} X_k\right).$$

Therefore, $m = \lim_{n\to\infty} P_{ij}(X_n) = P\left(\bigcap_{k=1}^{\infty} X_k\right) + \lim_{n\to\infty} \sum_{k=n}^{\infty} P_{ij}(X_k - X_{k+1}) = P_{ij}\left(\bigcap_{k=1}^{\infty} X_k\right) + O$. Observe that $M = \bigcap_{k=1}^{\infty} X_k \epsilon F$ is the required set.

LEMMA B. If F is a $\sigma$ — field of events over U and $P_{ij}: F \longrightarrow R$ is a real-valued bounded countably additive set function, then

$$P_{ij}(X) = P^{+}_{ij}(X) - P^{-}_{ij}(X)$$

for each $X \epsilon F$, where

$P^{+}_{ij}(X) = \sup [P_{ij}(Y) : X \supseteq Y \epsilon F]$ and

$P^{-}_{ij}(X) = -\inf [P_{ij}(Y) : X \supseteq Y \epsilon F]$ are both non-decreasing and countably additive set functions.

Proof. As in the proof of the previous Lemma, let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of subsets of X belonging to F such that $\lim_{n \to \infty} P_{ij}(X_n) = P^+_{ij}(X)$. Then for each n,

$$X_n \cup (X - X_n) = X$$

and hence $P_{ij}(X_n) + P_{ij}(X - X_n) = P_{ij}(X)$. Thus, $\lim_{n \to \infty}$

$P_{ij}(X - X_n) = P_{ij}(X) - \lim_{n \to \infty} P_{ij}(X_n) = P_{ij}(X) -$

$\sup [P_{ij}(Y): X \supseteq Y \in F] = \inf [P_{ij}(X) - P_{ij}(Y): X \supseteq Y \in F] = \inf [P_{ij}(X - Y): X \supseteq X - Y \in F] = \inf [P_{ij}(Z): X \supseteq Z \in F] = P^-_{ij}(X)$.

Therefore $P_{ij}(X) = \lim_{n \to \infty} P_{ij}(X) + \lim_{n \to \infty} P_{ij}(X - X_n) =$

$P^+_{ij}(X) - P^-_{ij}(X)$ for all $X \in F$.

If $X, Y \in F$ such that $X \subseteq Y$, then clearly $P^+_{ij}(X) \leqq P^+_{ij}(Y)$ and $P^-_{ij}(X) \leqq P^-_{ij}(Y)$ from the definition. These mean that $P^+_{ij}$ and $P^-_{ij}$ are non-decreasing set functions.

To show that $P^+_{ij}$ is countably additive, we first show finite additivity. Let $X, Y \in F$ such that $X \cap Y = \phi$. Then for each $X \cup Y \supseteq Z \in F$, we have $Z = (X \cap Z) \cup (Y \cap Z)$ so that $P_{ij}(Z) = P_{ij}(X \cap Z) + P_{ij}(Y \cap Z) \leqq P^+_{ij}(X) + P^+_{ij}(Y)$. Whence $P^+_{ij}(X \cup Y) \leqq P^+_{ij}(X) + P^+_{ij}(Y)$. By definition of $P^+_{ij}$ there exist sets $Z \supseteq X_n \in F$ and $Y \supseteq Y_n \in F$ such that $\lim P_{ij}(X_n) = P^+_{ij}(X)$ and $\lim P_{ij}(Y_n) = P^+_{ij}(Y)$. Thus for n big enough $X_n \cap Y_n = \phi$ and $P_{ij}(X_n \cup Y_n) = P_{ij}(X_n) + P_{ij}(Y_n)$. Hence $\lim P_{ij}(X_n \cup Y_n) = \lim P_{ij}(X_n) + \lim P_{ij}(Y_n) = P^+_{ij}(X) + P^+_{ij}(Y)$. Inasmuch as $X_n \cup Y_n \subseteq X \cup Y$, then $P^+_{ij}(X) + P^+_{ij}(Y) \leqq P^+_{ij}(X \cup Y)$.

Now, let $S = \bigcup_{i=1}^{\infty} S_i \in F$ where $S_1, S_2, \ldots, S_n, \ldots$ are pairwise disjoint sets also belonging to F. Then for $S \supseteq Z \in F$, $Z = \bigcup_{i=1}^{\infty} (S_i \cap Z)$ and $S_i \cap Z \subseteq S_i$ for all i. Hence $P_{ij}(Z) =$

$\sum_{i=1}^{\infty} P_{ij}(S_i \cap Z) \leqq \sum_{i=1}^{\infty} P^+_{ij}(S_i)$. This implies $P^+_{ij}(S) \leqq \sum_{i=1}^{\infty} P^+_{ij}(S_i)$. On the other hand $S_1 \cup S_2 \cup \ldots \cup S_n \subseteq S$ and since $P^+_{ij}$ is finitely additive and non-decreasing we have

$P^{+}_{ij}(S_1 \cup \ldots \cup S_n) = P^{+}_{ij}(S_1) + \ldots + P^{+}_{ij}(S_n) \leqq P^{+}_{ij}(S)$.

Whence $\sum_{i=1}^{\infty} P^{+}_{ij}(S_i) \leqq P^{+}_{ij}(S)$. The final result follows

THEOREM. If (U, F, P) is a matrix-valued probability space and $X \in F$ such that

$$P(X) = \begin{bmatrix} P_{11}(X) & P_{12}(X) & \ldots & P_{1n}(X) \\ P_{12}(X) & P_{22}(X) & \ldots & P_{2n}(X) \\ \cdots & \cdots & \cdots & \cdots \\ P_{1n}(X) & P_{2n}(X) & \ldots & P_{nn}(X) \end{bmatrix}$$

then

(i) $(U, F, P_{ii})$ for each $i = 1, 2, \ldots, n$ are ordinary real-valued probability spaces;

(ii) $P_{ij}$ $(i \neq j)$ for all $i, j = 1, 2, \ldots, n$ are finite countably additive set functions on F to [0,1] such that

$P_{ij} = P^{+}_{ij} - P^{-}_{ij}$, where $(U, F, P^{+}_{ij})$ and $(U, F, P^{-}_{ij})$ are ordinary real-valued probalility spaces.

Proof. Consider an arbitrary family $X_1, X_2, \ldots, X_n, \ldots$ of pairwise dispoint sets in F. Then

$(P_{ij}(\bigcup_{k=1}^{\infty} X_k)) = P(\bigcup_{k=1}^{\infty} X_k) = \sum_{k=1}^{\infty} P(X_k) = \sum_{k=1}^{\infty} (P_{ij}(S_k)) = (\sum_{k=1}^{\infty} P_{ij}(S_k))$.

Hence for all choices of i and j,

$$P_{ij}(\bigcup_{k=1}^{\infty} S_k) = \sum_{k=1}^{\infty} P_{ij}(S_k).$$

(i) This means that for each $X \in F$, $O \leqq P(X) = (P_{ij}(X)) \leqq I = (d_{ij})$ where $d_{ij} = O$ for $i \neq j$ and $d_{ii} = 1$. Thus, $O \leqq P_{ii}(X)$ and $(d_{ij} - P_{ij}(X)) \geqq O$ so that $1 - P_{ii}(X) \leqq O$ or $P_{ii}(X) \leqq 1$. Whence $(U, F, P_{ii})$ is an ordinary probability space for all $i = 1, 2, \ldots, n$.

(ii) For all $i \neq j$, note that since $(P_{ij}(X))$ is positive semi-definite, all its principal minors must be non-negative, that is to say,

$$\begin{vmatrix} P_{ii}(X) & P_{ij}(X) \\ P_{ij}(X) & P_{jj}(X) \end{vmatrix} = P_{ii}(X)\, P_{jj}(X) - P^2_{ij}(X) \geqq O.$$

This implies that $P^2_{ij}(X) \leqq P_{ii}(X)\, P_{jj}(X) \leqq 1.1 = 1$ or $-1 \leqq P_{ij}(X) \leqq +1$.

By Lemma A, there exists a set $M \epsilon F$ such that $P_{ij}(M)$ (which is less than or equal to 1) is maximum. For each $X \epsilon F$, set $X_1 = X \cap M$ and $X_2 = X - X_1 = X - M$.

Then $P^+_{ij}(X_2) = O$. For, suppose not, that is $P^+_{ij}(X_2) > O$. Then by the definition of the sup there exists a set $Y \subseteq S_2$ with $Y \epsilon F$ such that $P_{ij}(Y) > O$. Since $Y \cap M \subseteq S_2 \cap M = \phi$, then $Y \cup M \epsilon F$ and $P_{ij}(Y \cup M) = P_{ij}(Y) + P_{ij}(M) > P_{ij}(M)$, contrary to the maximality of $P_{ij}(M)$.

Similarly, $P^-_{ij}(X_1) = O$, for, if not $P^-_{ij}(X_1) > O$, then by definition of the inf there exists a set $Z \subseteq X_1$ with $Z \epsilon F$ such that $-P_{ij}(Z) > O$ or $P_{ij}(Z) < O$. Since $Z \subseteqq X_1 \subseteqq M$, then $(M - Z) \cup Z = M$ and

$$P_{ij}(M - Z) + P_{ij}(Z) = P_{ij}(M) \text{ or}$$
$$P_{ij}(M - Z) = P_{ij}(M) - P_{ij}(Z) > P_{ij}(M),$$

again contrary to the maximality of $P_{ij}(M)$.

From the conclusions $P^-_{ij}(X_1) = O$ and $P^+_{ij}(X_2) = O$ of the two previous paragraphs, it follows then that

$P^+_{ij}(X) = P^+_{ij}(X_1 \cup X_2) = P^+_{ij}(X_1) + P^+_{ij}(X_2) = P^+_{ij}(X_1)$ and $P^-_{ij}(X) = P^-_{ij}(X_1 \cup X_2) = P^-_{ij}(X_1) + P^-_{ij}(X_2) = P^-_{ij}(X_2)$.

From these it follows that

$P_{ij}(X_1) = P^+_{ij}(X_1) - P^-_{ij}(X_1) = P^+_{ij}(X_1) = P^+_{ij}(X)$ and $P_{ij}(X_2) = P^+_{ij}(X_2) - P^-_{ij}(X_2) = -P^-_{ij}(X)$.

Therefore, for an arbitrary $X \epsilon F$, we have

$$O \leqq P^+_{ij}(X) = P_{ij}(X_1) = |\, P_{ij}(X_1) \,| \leqq 1$$

and

$$O \leqq P^-_{ij}(X) = -P_{ij}(X_2) = |\, P_{ij}(X_2) \,| \leqq 1.$$

These relations complete the proof that $P^+_{ij}$ and $P^-_{ij}$ for all $i \neq j$ are probability functions and $P_{ij} = P^+_{ij} - P^-_{ij}$.

To a mathematically trained reader, it is now almost obvious that numerous standard results in ordinary real-valued probability do extend to the case of matrix-valued probability. For a complete exposition of these results please refer to the monograph of the author which will be published by the Bureau of the Census and Statistics[4].

3. **Conditional Probability and Independence.** For purposes of illustration we shall here develop the notion of conditional probability and prove the Bayes Theorem in matrix-valued probability spaces.

DEFINITION 3. Events $X_1, X_2, \ldots, X_n \in F$ in a matrix-valued probability space $(U, F, P)$ are said to be independent if and only if for any subset $[Y_1, \ldots, Y_m]$ of $[X_1, \ldots, X_n]$, $P(Y_1 \cap \ldots \cap Y_m) = P(Y_1) \ldots P(Y_m)$.

Observe that this definition implies that the product appearing on the right side of the above equality is not only defined but is also positive semi-definite and hence their factors commute with one another. (Recall that two positive semi-definite matrices have a positive semi-definite product if and only if they commute.)

DEFINITION 4. If $X, Y \in F$ of a matrix-valued probability space $(U, F, P)$ and $P(X)$ is non-singular (i. e. positive definite) and commutes with $P(X \cap Y)$, then the conditional probability of $Y$ given $X$ is defined by

$$P(Y \mid X) = P(Y \cap X)\, P(X)^{-1}.$$

From Definition 4, note that if $P(X)$ commutes with $P(Y \cap X)$, then $P(X)^{-1}$ also commutes with $P(Y \cap X)$ and $P(Y \mid X)$ is well-defined.

PROPOSITION. Let $(U, F, P)$ be a matrix-valued probability space. Then

(1) for all $Y, X \in F$ such that $X \subseteq Y$ we have $P(Y \mid X) = I$. In particular, $P(X \mid X) = I$;

(2) if for all $X \epsilon F$ and every family $Y_1, Y_2, \ldots, Y_n, \ldots$ of pairwise disjoint subsets in F the probabilities $P(Y_i \mid X)$ are defined for each $i = 1, 2, \ldots, n, \ldots,$ then

$$P\left(\bigcup_{i=1}^{\infty} Y_i \mid X\right) = \sum_{i=1}^{\infty} P(Y_i \mid X)$$

(3) $X, Y \epsilon F$ are independent if and only if $P(Y \mid X) = P(Y)$;

(4) if $\bigcup_{i=1}^{\infty} X_i = U$ and $P(Y \mid X_i)$ are well-definied and $X_i \cap X_j \cap Y = \phi$ $(i \neq j)$ for all i and j, then

$$P(Y) = \sum_{i=1}^{\infty} P(Y \mid X_i) P(X_i).$$

Proof. (1) If $X \subseteq Y$, then certainly $P(Y \cap X) = P(X)$ commutes with $P(X)^{-1}$ and $P(Y \mid X) = P(Y \cap X) P(X)^{-1} = I$.

(2) If $P(Y_i \mid X) = P(Y_i \cap X) P(X)^{-1}$ for all $i = 1, 2, \ldots, n, \ldots,$ then remembering that $(Y_i \cap X) \cap (Y_j \cap X) = Y_i \cap Y_j \cap X = \phi$ for all $i \neq j$, then $\sum_{i=1}^{\infty} P(Y_i \mid X) = \sum_{i=1}^{\infty} P(Y_i \cap X) P(X)^{-1} = \left(\sum_{i=1}^{\infty} P(Y_i \cap X)\right) P(X)^{-1}$ $= P\left(\bigcup_{i=1}^{\infty}(Y_i \cap X)\right) P(X)^{-1} = P\left(\left(\bigcup_{i=1}^{\infty} Y_i\right) \cap X\right) P(X)^{-1}$ $= P\left(\bigcup_{i=1}^{\infty} Y_i \mid X\right).$

(3) If Y and X are independent, so that $P(Y \cap X) = P(Y) P(X)$, then $P(Y) = P(Y \cap X)P(X)^{-1} = P(Y \mid X)$. Conversely, if $P(Y) = P(Y \mid X)$, then $P(Y) = P(Y \cap X)P(X)^{-1}$ and therefore $P(Y \cap X) = P(Y)P(X)$.

(4) By hypothesis $P(Y \mid X_i) = P(Y \cap X_i)P(X_i)^{-1}$ or $P(Y \mid X_i)P(X_i) = P(Y \cap X_i)$ for all $i = 1, 2, \ldots, n, \ldots$ Then since $Y_i \cap Y_j \cap X = \phi$ for $i \neq j$,

then

$$\sum_{i=1}^{\infty} P(Y \mid X_i)P(X_i) = \sum_{i=1}^{\infty} P(Y \cap X_j) = P(\bigcup_{i=1}^{\infty} (Y \cap X_j)) = P(Y \cap \bigcup_{i=1}^{\infty} X_i) = P(Y \cap U) = P(Y).$$

BAYES THEOREM. Under the hypothesis of (4) of the previous proposition and if $P(X_k \mid Y)$ are defined, then

$$P(X_k \mid Y) \sum_{i=1}^{\infty} P(Y \mid X_i)P(X_i) = P(Y \mid X_k)P(X_k).$$

Proof. By hypothesis $P(X_k \mid Y) = P(X_k \cap Y)P(Y)^{-1}$ and $P(Y \mid X_k) = P(Y \cap X_k)P(X_k)^{-1}$. Thus, $P(x_k \mid Y)P(Y) = P(X_k \cap Y) = P(Y \mid X_k)P(X_k)$. By substituting the result of (4) in the previous proposition, Bayes theorem is thus obtained.

REMARKS. A particular type of matrix-valued probability space (U, F, P) for which $P(Y \mid X)$ is invariably defined for every $X, Y \epsilon F$ is one in which

$$P(X) = \begin{pmatrix} P_1(X) & P_1(X) - P_2(X) \\ P_1(X) - P_2(X) & P_2(X) \end{pmatrix}$$

where $P_1$ and $P_2$ are any two ordinary probailities defined on (U, F).

In this case observe that the product of any positive semi-definite matrices of the above form is always positive semi-definite, since any two of them commute.

## REFERENCES

[1] KAWADA, Yukiyosi: "Uber eine verbandstheoretische Begründung der Wahrscheinlichkeitsrechnung", *Japanese Journal of Mathematics,* 18 (1943), pp. 887-972.

[2] KOLMOGOROV, A.N.: *Foundations of the Theory of Probability.* 2nd English Edition. New York: Chelsea Publishing Company, 1956.

[3] LOS, JERZY: "On the axiomatic treatment of probability", *Colloquium Mathematicum,* 3 (1955), pp. 125-137.

[4] SIOSON, F.M.: *The Theory of Operator-Valued Probabilities.* A research monograph to be published by the Bureau of Census and Statistics, Manila.