Conditional Probability

on a σ-Complete Boolean Algebras

by

Ardel J. Boes

Supported by the Air Force Office of Scientific Research

Contract AFOSR 955-65

Glen Baxter, Project Director

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series No. 85

August 9, 1966

INTRODUCTION

There are various models for probability. Each axiomatic treatment of probability seems to be concerned with the decisions of what events should be and what kind of function of events a probability should be. Two models will be discussed here.

The almost universally accepted axiomatic of the calculus of probability is that of Kolmogorov [14], published in 1933. The model is a triple $(X, \mathcal{A}, \lambda)$, where X is an abstract space, $\mathcal{A}$ is a σ-field of subsets of X, and λ is totally finite σ-additive measure defined on $\mathcal{A}$ with the property that $\lambda(X) = 1$. The Kolmogorov model is not without defects, Kolmogorov has criticized it himself (v. [15]). Various examples have been published which indicate there is a certain amount of pathology inherent in the Kolmogorov model (v. [4], [1], and [10]). The example of Dieudonne [4] is related to the present work. His example is a probability space $(X, \mathcal{A}, \lambda)$ and a σ-subfield $\mathcal{B}$ of $\mathcal{A}$ for which there is no function $\lambda(\cdot,\cdot)$, defined on $\mathcal{A} \times X$ which for fixed A in $\mathcal{A}$ is a $\mathcal{B}$-measurable function of x, and for which $\lambda(A \cap B) = \int_B \lambda(A,x) d\, \lambda_{|\mathcal{B}}(x)$ for each B in $\mathcal{B}$; and for fixed x in X, a probability measure on $\mathcal{A}$. That is there is no regular conditional probability $\lambda(\cdot,\cdot | \mathcal{A}, \mathcal{B})$ on $\mathcal{A} \times X$ (v. Jirina [11]). The notion of regular conditional probability on $\mathcal{A} \times X$ is introduced in chapter 2.

Much work has been done to eliminate this difficulty within the Kolmogorov framework by making the model less general. In this connection the concepts of compact measure and perfect measure are important. λ is perfect in case for every real-valued, $\mathcal{Q}$-measurable function f on X and for every set of real numbers A for which $f^{-1}(A)$ is in $\mathcal{Q}$, there is a Borel set B contained in A for which $\lambda(f^{-1}(B)) = \lambda(f^{-1}(A))$. A class $\mathcal{C}$ of subsets of a set X is compact, if for each sequence C_n in $\mathcal{C}$ the relation $\bigcap_{i=1}^{\infty} C_i \neq \phi$ for $n=1,2,\ldots$ implies $\bigcap_{i=1}^{\infty} C_i \neq \phi$. A finitely additive probability measure μ defined on a field $\mathcal{B}$ is compact if there is a compact class $\mathcal{C}$ which approximates $\mathcal{B}$ with respect to μ, that is for each A in $\mathcal{B}$ and $\eta > 0$ there is a set C in $\mathcal{C}$ and a set B in $\mathcal{B}$ such that $B \subset C \subset A$ and $\mu(A-B) < \eta$. If $(X, \mathcal{Q}, \lambda)$ is a Kolmogorov model such that $\mathcal{Q}$ is countably generated and λ is a compact measure or a perfect measure, a regular conditional probability on $\mathcal{Q} \times X$ always exists (v. Jiřina [11]). There is a close connection between compact measures and perfect measures (v. [20] and [21]).

Another objection to the Kolmogorov model points out that it does not admit the identification of almost identical events, or stated differently, it does not allow the introduction of a strictly positive probability measure. Actually, the pathology in the area of conditioning seems to arise from this difficulty. Roughly speaking, the sets of probability zero can add up to a set which is too big.

In order to avoid the latter difficulty it may be supposed that the events form a Boolean algebra and that the probability is strictly positive. A Boolean algebra is a ring with unit in which every element is idempotent. A Boolean algebra is σ-complete if every sequence of elements of the Boolean algebra has a supremum and an infimum in the Boolean algebra. A Boolean algebra is complete if every subset of the algebra has a supremum and infimum in the algebra. Then for every Boolean algebra B_o with a strictly positive, finitely additive measure λ_o there exists a unique (up to an isomorphism) σ-complete Boolean algebra B with a strictly positive, countably additive measure λ such that B is an extension of B_o, λ is an extension of λ_o, and B is the smallest σ-complete subalgebra of B containing B_o. Most notable in this area are the works of Kappos, particularly [12] in which he undertakes to study the structure of probability on a Boolean algebra. The lack of a basic space of which the events are subsets causes difficulty in defining some probabilistic notions. The notion of random variable and its expectation has been defined by Olmsted [19] and others, however none of these developments has gone far enough, namely in the area of conditioning, to recognize the merit of the Boolean approach. A Kolmogorov model can be converted to the Boolean approach by forming the quotient of the σ-field of measurable events modulo the null events, thus circumventing many difficult measurability problems.

On the other hand, a σ-complete Boolean algebra with a probability defined on it is isomorphic to the σ-field generated by the open-closed sets of the Stone space of the Boolean algebra modulo the null sets (the Loomis representation theorem). Of course, the Stone space of a Boolean algebra is a compact topological space and if the Boolean algebra is the quotient of a σ-field of measurable events modulo the null events of a

Kolmogorov model, theoretically the integration is largely equivalent on the Stone space, but the apparent impossibility of describing it explicitly and the necessity of carrying an isomorphism throughout a problem limits its usefulness [22]. This leads to the considerations of the assumption that the basic space is a locally compact topological space (v. Bourbaki [3], especially Tulcea [8]). While this approach is quite adequate for the treatment of integration in geometrical types of spaces, it is out of place in probability theory [22].

In this work the problem of conditioning is considered on a Boolean algebra. Such a treatment is conspicuously absent from the works of Kappos which is the prime motivation of the present study. Apparently this topic has been avoided because of the Loomis representation theorem, but as was mentioned above this approach has its drawbacks in general considerations. In the study of conditioning on the Stone space, the intrinsic properties of the space come into play and to the probabilist the work becomes cumbersome. The natural alternative to this is to carry out the theory on the Boolean algebra. With this in mind it becomes apparent a generalized notion of probability and integral are needed for the study. These come quite naturally and are developed in chapter 2 where conditional probability on a Boolean algebra is also introduced. This theory follows smoothly and the objections to the Stone space approach do not apply. Moreover, there is no lack of regularity as in the Kolmogorov model. Doob [5] has shown that if a regular conditional probability relative to a σ-field of sets exists, then the conditional expectation of an integrable function is given by the integral of the function with respect to this conditional distribution (making the obvious notational conventions). The analogue to this theorem holds in general in the Boolean model and is quite easily proved. The lack of numerical values associated with random variables, probabilities (as defined here), and conditional probabilities is only apparent as is shown in this work.

1. CONDITIONING ON THE STONE SPACE OF A BOOLEAN ALGEBRA

1.1. Preliminaries

Throughout this work a Boolean algebra (σ-complete Boolean algebra) whose elements are sets will be denoted by field (σ-field). Let the events which are to be probable form a σ-complete Boolean algebra, then there is isomorphic to it a perfect reduced field of subsets of a space by the Stone representation theorem. (A field of subsets of a space is reduced in case any two different points are separated by a set in this field. A field of subsets of a space is perfect if every maximal filter of the field is determined by a point of the space. A filter is determined by a point if the filter is the class of all subsets of the field which contain the point.) For a perfect reduced field of subsets of a space, a topology can be defined in the space so that the space becomes a compact totally disconnected (Hausdorff) space and the field becomes the class of open-closed subsets of the topological space. See Sikorski [23] for the definitions of the above terms and the proofs of the remarks. For a treatment of the topological notions used in this work see for example Kelley [13]. Thus, if A is a σ-complete Boolean algebra, there is an isomorphism Θ such that $\Theta(A) = \mathcal{E}$ is the class of open-closed subsets of a perfect reduced field of subsets of a space Y. Y is called the Stone space of A. $\mathcal{E}$ forms a σ-complete Boolean algebra but only finite supremum and infimum correspond to set-theoretic union and intersection, respectively. Let $\mathcal{B}(\mathcal{E})$ be the σ-field generated by $\mathcal{E}$, that is the smallest σ-field of subsets of Y containing $\mathcal{E}$. Every countable union of sets of $\mathcal{E}$ is contained in set of $\mathcal{E}$ but need not be a set of $\mathcal{E}$. The following conventions will be made throughout this work, unless specifically noted otherwise: the symbols $\vee$ and $\wedge$ will mean the algebraic supremum and infimum, respectively; and the symbols $\cup$ and $\cap$ will mean set-theoretic union and intersection, respectively. The remarks that follow are motivated by the Aumann proof of the Loomis representation theorem which can be found in Kappos [12]. Let $D = (\bigvee_{n=1}^{\infty} E_n) - (\bigcup_{n=1}^{\infty} E_n)$, $\{E_n\}$ a sequence of open-closed sets in Y, and let Δ be the σ-ideal in $\mathcal{B}(\mathcal{E})$ generated by the totality of all such D, then $\mathcal{E}$ is isomorphic to $\mathcal{B}(\mathcal{E})/\Delta$. Each $B \in \mathcal{B}(\mathcal{E})$ can be represented uniquely by

$$B = E \dotplus I \quad (\dotplus \text{ is symmetric difference, the Boolean } +)$$

where E is an open-closed set and I belongs to Δ.

If λ is a strictly positive probability defined on A, define μ on $\mathcal{E}$ by

$$\mu(E) = \lambda(\theta^{-1}(E)), \quad E \in \mathcal{E}.$$

In view of the isomorphism θ, μ is a finitely additive, strictly positive probability on $\mathcal{E}$. Define μ on $\mathcal{B}(\mathcal{E})$ by

$$\mu(B) = \mu(E),$$

where E is the uniquely determined open-closed set of the representation given above. Then $\mu(I) = 0$ for every $I \in \Delta$. The following theorems, 1.1 and 1.2, are known, but it is convenient in the present work to have constructive proofs.

<u>Theorem 1.1.</u> The set function μ defined on $\mathcal{B}(\mathcal{E})$ is a probability measure on $\mathcal{B}(\mathcal{E})$.

<u>Proof.</u> Let $\{B_n\}$ be a sequence of $\mathcal{B}(\mathcal{E})$, then

$$\mu(\bigcup_{n=1}^{\infty} B_n) = \mu(\bigcup_{n=1}^{\infty}(E_n \dot{-} I_n)) \leq \mu((\bigcup_{n=1}^{\infty} E_n) \cup (\bigcup_{n=1}^{\infty} I_n)) = \mu(\bigvee_{n=1}^{\infty} E_n), \text{ and}$$

$$\mu(\bigcup_{n=1}^{\infty} B_n) \geq \mu(\bigcup_{n=1}^{\infty}(E_n - I_n)) \geq \mu((\bigcup_{n=1}^{\infty} E_n) - (\bigcup_{n=1}^{\infty} I_n)) = \mu(\bigvee_{n=1}^{\infty} E_n). \text{ Moreover,}$$

$$\mu(\bigvee_{n=1}^{\infty} E_n) = \lambda(\theta^{-1}(\bigvee_{n=1}^{\infty} E_n)) = \lambda(\bigvee_{n=1}^{\infty} \theta^{-1}(E_n)) = \lim_{n\to\infty} \lambda(\bigvee_{i=1}^{n} \theta^{-1}(E_i))$$

$$= \lim_{n\to\infty} \lambda(\theta^{-1}(\bigvee_{i=1}^{n} E_i)) = \lim_{n\to\infty} \mu(\bigvee_{i=1}^{n} E_i) = \lim_{n\to\infty} \mu(\bigcup_{i=1}^{n} E_i).$$

If B_1, B_2 are disjoint elements of $\mathcal{B}(\mathcal{E})$, $B_1 = E_1 \dot{-} I_1$ and $B_2 = E_2 \dot{-} I_2$, then

$$\mu(B_1 \cup B_2) = \mu(E_1 \cup E_2).$$

Since $\mu(E) > 0$, for all E in $\mathcal{E}$ for which $E \neq \phi$, and $E_1 \cap E_2 \subset I_1 \cup I_2$, it follows that $E_1 \cap E_2 = \phi$. Hence

$$\mu(B_1 \cup B_2) = \mu(E_1) + \mu(E_2) = \mu(B_1) + \mu(B_2).$$

Now let $\{B_n\}$ be a sequence of pairwise disjoint elements of $\mathcal{B}(\mathcal{E})$, then by the above considerations it is clear that

$$\mu(\bigcup_{n=1}^{\infty} B_n) = \mu(\bigvee_{n=1}^{\infty} E_n) = \lim_{n\to\infty} \mu(\bigcup_{i=1}^{n} E_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(E_i) = \sum_{n=1}^{\infty} \mu(E_n) = \sum_{n=1}^{\infty} \mu(B_n).$$

<u>Theorem 1.2</u>. The elements of the σ-ideal Δ are sets of the first category in Y.

<u>Proof</u>. If $\{E_n\}$ is a sequence of open-closed sets in Y, $\bigvee_{n=1}^{\infty} E_n$ is the smallest open-closed set containing $\bigcup_{n=1}^{\infty} E_n$. Then $D = (\bigvee_{n=1}^{\infty} E_n) - (\bigcup_{n=1}^{\infty} E_n)$ is closed, since D is the intersection of the closed sets $\bigvee_{n=1}^{\infty} E_n$ and $(\bigcup_{n=1}^{\infty} E_n)^c$, where A^c means the complement of A.

Suppose there is a closed set C such that $\bigvee_{n=1}^{\infty} E_n \supset C \supset \bigcup_{n=1}^{\infty} E_n$, then $C = \bigcap_{\alpha\in I} E_\alpha$, where E_α is an open-closed set in Y for every α in I, where I is some index set. But then, since $E_\alpha \supset \bigcap_{\alpha\in I} E_\alpha \supset \bigcup_{n=1}^{\infty} E_n$, all α in I, it follows that for every α in I, $E_\alpha \supset E_n$, $n=1,2,\ldots$. Thus $E_\alpha \supset \bigvee_{n=1}^{\infty} E_n$, X in I, because $\bigvee_{n=1}^{\infty} E_n$ is the smallest open-closed set containing each E_n, $n=1,2,\ldots$. Therefore $C = \bigcap_{\alpha\in I} E_\alpha \supset \bigvee_{n=1}^{\infty} E_n$, that is closure $(\bigcup_{n=1}^{\infty} E_n) = \bigvee_{n=1}^{\infty} E_n$. From this it follows that (A^o means the interior of A) $(\text{closure }(D))^o = D^o$

$$= (\bigvee_{n=1}^{\infty} E_n)^o \cap ((\bigcup_{n=1}^{\infty} E_n)^c)^o = \bigvee_{n=1}^{\infty} E_n \cap (\text{closure }(\bigcup_{n=1}^{\infty} E_n))^c = \phi.$$

The σ-ideal generated by sets of the form D is the set of all $B \subset D_1 \cup \ldots \cup D_n \cup \ldots$, where D_i, $i=1,2,\ldots$ is a set of the form

D. By above considerations the sets of the form D are nowhere dense in Y, hence the sets of the σ-ideal are of the first category in Y.

The following result will be used later and a direct proof is included here for the sake of completeness.

<u>Theorem 1.3</u>. If the continuous (real-valued) functions f and g defined on Y are equal almost everywhere, they are equal everywhere.

<u>Proof</u>. If $f\colon Y \to R$ is continuous, then for each net $\{S_n, n \in D\}$ in Y which converges to a point y, the net $\{f(S_n), n \in D\}$ converges to $f(y)$. Let $g\colon Y \to R$ be continuous and suppose $f(y) = g(y)$, for all y in N^c with $\mu(N) = 0$. By theorem 1.2, N is a set of the first category in Y. By the Baire category theorem N^c is dense in Y, thus there is a net in N^c converging to y for every y in N. If $\{S_n, n \in D\}$ is a net in N^c converging to y, y in N, the net $\{f(S_n), n \in D\} = \{g(S_n), n \in D\}$ converges to $g(y)$. Since R is Hausdorff, $f(y) = g(y)$.

1.2. <u>Conditioning</u>

Turn now to the study of conditional probability for the probability space $(Y, \mathcal{B}(\mathcal{E}), \mu)$. The sets of $\mathcal{B}(\mathcal{E})$ are called Baire sets and functions on Y which are $\mathcal{B}(\mathcal{E})$-measurable are called Baire functions. Let $\mathcal{G}$ be a full σ-subfield of $\mathcal{B}(\mathcal{E})$; ''full'' means that for each J in $\mathcal{G}$, the open-closed set of $\mathcal{E}$ which is μ-equivalent to J is also in $\mathcal{G}$. Moreover, assume that $\mu|_{\mathcal{G}}$, the restriction of μ to $\mathcal{G}$ is complete. By the Radon-Nikodyn theorem, define $p(E,\cdot|\mathcal{B}(\mathcal{E}),\mathcal{G}) = p(E,\cdot)$ as any $\mathcal{G}$-measurable function for which $\mu(E \cap J) = \int_J p(E,\cdot)\, d\mu|_{\mathcal{G}}(y)$

holds for every J in $\mathcal{J}$, for each E in $\mathcal{E}$. Since $\mu|\mathcal{J}$ is complete it is possible to choose a version $p(E,\cdot)$ so that $0 \leq p(E,\cdot) \leq 1$. Define $s_n(E,\cdot)$ by

$$s_n(E,y) = \sum_{k=1}^{2^n+1} \frac{k-1}{2^n} \chi_{A_{n,k}}(y),$$

where $A_{n,k} = \{y: \frac{k-1}{2^n} \leq p(E,y) < \frac{k}{2^n}\}$, $k = 1,\ldots,2^n+1$, for every natural number n. Then $s_n(E,\cdot)$ converges uniformly to $p(E,\cdot)$. For each $A_{n,k}$ there is $E_{n,k}$ in $\mathcal{E}$ such that $\mu(A_{n,k} \dotdiv E_{n,k}) = 0$. As was seen in the proof of theorem 1.1, the $E_{n,k}$ are disjoint for fixed n because the $A_{n,k}$ are disjoint for fixed n. Since $Y = \bigcup_{k=1}^{2^n+1} A_{n,k}$, $\mu(Y - \bigcup_{k=1}^{2^n+1} E_{n,k}) = 0$. But $Y - \bigcup_{k=1}^{2^n+1} E_{n,k}$ is an open-closed set so $Y = \bigcup_{k=1}^{2^n+1} E_{n,k}$. If for some k, $1 \leq k \leq 2^n+1$, $A_{n+1,j} \cup A_{n+1,i} = A_{n,k}$, then $E_{n+1,j} \cup E_{n+1,i} = E_{n,k}$, where $1 \leq i, j \leq 2^{n+1}+1$.

Let $c_n(E,y) = \Sigma_{k=1}^{2^n+1} \frac{k-1}{2^n} \chi_{E_{n,k}}(y)$, then $c_n(E,\cdot)$ is continuous and if $c(E,y) = \lim_n c_n(E,y)$ for all y, then $c_n(E,\cdot)$ converges uniformly to $c(E,\cdot)$ so that $c(E,\cdot)$ is a continuous function. Since $\mathcal{J}$ is full, the sets $E_{n,k}$ are in $\mathcal{J}$ which implies that $c(E,\cdot)$ is $\mathcal{J}$-measurable as the limit of a sequence of $\mathcal{J}$-measurable functions. Moreover, there is N in $\mathcal{J}$, $\mu(N) = 0$, such that $c_n(E,y) = s_n(E,y)$

for y not in N, all $n=1,2,\ldots$. Hence $c(E,\cdot)$ is a continuous version of the functions defined above by the Radon-Nikodym theorem.

In view of the properties of conditional probabilities as defined in Doob [5], there is N in $\mathcal{G}$, $\mu(N) = 0$, for which $c(E_1,y) + c(E_2,y) = c(E_1 \cup E_2,y)$ if y is not in N and E_1 and E_2 are disjoint open-closed subsets of Y. By theorem 1.3, it is clear that for fixed y in Y, the function $c(\cdot,y)$ is a finitely additive set function on $\mathcal{E}$. Then by further use of the properties of conditional probabilities, $c(\cdot,y)$ is a finitely additive probability on $\mathcal{E}$ for fixed y. But a finitely additive probability on a perfect field of sets can be extended to a probability on the σ-field generated by the perfect field (v. Sikorski [23]). There will be no confusion if we denote the extension of $c(\cdot,y)$ to $\mathcal{B}(\mathcal{E})$ by $c(\cdot,y)$.

Let $\mathcal{B}$ denote the class of all B in $\mathcal{B}(\mathcal{E})$ for which $c(B,\cdot)$ is $\mathcal{G}$-measurable and for which

$$\mu(B \cap J) = \int_J c(B,\cdot)\, d\, \mu|_{\mathcal{G}}\,(y)$$

holds for all J in $\mathcal{G}$. It is obvious that $\mathcal{B} \supset \mathcal{E}$ and that $\mathcal{B}$ is a monotone class. By virtue of theorem 6B [6] $\mathcal{B}(\mathcal{E}) \subset \mathcal{B}$, that is $\mathcal{B}(\mathcal{E}) = \mathcal{B}$.

The above results can be summarized in the following theorem.

<u>Theorem 1.4</u>. The function $c(\cdot,\cdot|\mathcal{B}(\mathcal{E}),\mathcal{G}) = c(\cdot,\cdot)$ is a regular conditional probability distribution on $\mathcal{B}(\mathcal{E}) \times Y$.

For any integrable function f on Y, by the Radon-Nikodym theorem let $E(f,\cdot)$ be any $\mathcal{G}$-measurable function for which

$$\int_J f \, d\mu(y) = \int_J E(f,\cdot) \, d\mu|_{\mathcal{G}} (y)$$

holds for all J in $\mathcal{G}$. For set characteristic functions χ_B on Y, we may take $E(\chi_B,\cdot) = c(B,\cdot)$ since

$$\int_J \chi_B \, d\mu(y) = \mu(B \cap J) = \int_J E(\chi_B,\cdot) \, d\mu|_{\mathcal{G}} (y)$$

holds for all J in $\mathcal{G}$. If $f = \sum_{i=1}^{n} \alpha_i \chi_{B_i}$, we may take

$E(f,\cdot) = \sum_{i=1}^{n} \alpha_i \, E(\chi_{B_i},\cdot)$ since

$$\int_J f \, d\mu(y) = \sum_{i=1}^{n} \alpha_i \, \mu(B_i \cap J) = \sum_{i=1}^{n} \alpha_i \int_J E(\chi_{B_i},\cdot) \, d\mu|_{\mathcal{G}} (y) = \int E(f,\cdot) d\mu|_{\mathcal{G}} (y)$$

holds for all J in $\mathcal{G}$. If f is any function on Y whose expectation exists, let f_n be a sequence of simple functions converging to f. By the properties of conditional expectation as defined in Doob [5], $\lim_n E(f_n,\cdot)$ exists. Let $E(f,\cdot) = \lim_n E(f_n,\cdot)$, then $\mu|_{\mathcal{G}}$ complete implies that $E(f,\cdot)$ is a $\mathcal{G}$-measurable function for which

$$\int_J f d\mu(y) = \lim_n \int_J f_n d\mu(y) = \lim_n \int_J E(f_n,\cdot) \, d\mu|_{\mathcal{G}} (y) = \int_J E(f,\cdot) d\mu|_{\mathcal{G}} (y)$$

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

If the $E(f,\cdot)$ are chosen as above, the following result can be obtained.

Theorem 1.5. If f is an integrable function on Y, $E(f,\cdot) = \int_Y f\, c(dy,\cdot)$.

Proof. If B is a Baire set,

$$E(\chi_B,\cdot) = c(B,\cdot) = \int_B c(dy,\cdot) = \int_Y \chi_B\, c(dy,\cdot)\ .$$

If $f = \sum_{i=1}^{n} \alpha_i \chi_{B_i}$,

$$E(f,\cdot) = \sum_{i=1}^{n} \alpha_i E(\chi_{B_i},\cdot) = \sum_{i=1}^{n} \alpha_i \int_Y \chi_{B_i}\, c(dy,\cdot) = \int_Y f\, c(dy,\cdot).$$

If f is an integrable function, let f_n be a sequence of simple functions converging to f. Then

$$E(f,\cdot) = \lim_n E(f_n,\cdot) = \lim_n \int_Y f_n\, c(dy,\cdot) = \int_Y f\, c(dy,\cdot).$$

For this work we supposed the subfield $\mathcal{G}$ to be full and complete in $(Y, \mathcal{B}(\mathcal{E}),\mu)$. The assumption of completeness is a minor one for probability theory since we can take $\mathcal{G}$ to be the completion of the subfield. A discussion of how much the supposition that $\mathcal{G}$ be full restricts the problem will be included in the next section.

1.3. Relation to Probability Spaces

If $(X, \mathcal{A}, \lambda)$ is a probability space, then $\mathcal{A}/\lambda$ ($\mathcal{A}$ modulo the σ-ideal of λ-null sets of $\mathcal{A}$) is a complete Boolean algebra. Let $(Y, \mathcal{B}(\mathcal{E}), \mu)$ be the probability space formed above where Y is the Stone space of $\mathcal{A}/\lambda$. The relation between real-valued functions on the two spaces is in Dieudonne [4] and Halmos [7]. We have already presented part of this in Theorem 1.3. This relation is summarized in the following paragraph.

For every class of sets A_i in $\mathcal{A}$ and real numbers α_i, $i=1,\dots,n$, define

$$\theta\Big(\sum_{i=1}^{n} \alpha_i \, \chi_{A_i}\Big) = \sum_{i=1}^{n} \alpha_i \, \chi_{\theta([A_i])},$$

where $[A_i]$ is the residue class to which A_i belongs. It is clear that this is a continuous function on Y, that is simple functions on X go into continuous functions on Y which take on only a finite number of values. If f is a bounded $\mathcal{A}$-measurable function, there is a sequence s_n of simple functions converging to f. Define

$$\theta(f) = \lim_n \theta(s_n),$$

then

$$\theta(f+g) = \lim_n \theta(s_n + t_n) = \theta(f) + \theta(g).$$

Moreover, $||f||_\infty = ||\theta(f)||$, thus for every $\varepsilon > 0$, there is N such that $n > N$ implies

$$||\theta(s_n) - \lim_n \theta(s_n)|| = ||s_n - f||_\infty < \varepsilon \quad ,$$

that is $\lim_n \theta(s_n)$ is a continuous function on Y.

Let $\mathcal{A}_o \subset \mathcal{A}$ be a σ-subfield of $\mathcal{A}$ and define $\mathcal{A}_o/\lambda$ to be the set of all residue classes $[A]$ in $\mathcal{A}/\lambda$ for which there is A_o in $\mathcal{A}_o$ such that A_o is in $[A]$. This is a complete Boolean subalgebra of $\mathcal{A}/\lambda$. If θ is the mapping referred to in section 1.1, $\theta(\mathcal{A}_o/\lambda)$ is a class of open-closed sets which forms a field of sets. Let $\mathcal{E}_o$ denote this class and let $\mathcal{B}(\mathcal{E}_o)$ be the σ-field generated by $\mathcal{E}_o$. Then $\mathcal{B}(\mathcal{E}_o) \subset \mathcal{B}(\mathcal{E})$ so μ is defined on $\mathcal{B}(\mathcal{E}_o)$. From this it is seen that $\mathcal{B}(\mathcal{E}_o)$ is full.

In view of the results of theorems 1.4 and 1.5, from the probability theoretic standpoint it is advantageous to study conditional probability on the Stone space of a probability space. However, the student of probability is taken far afield to attain theorem 1.4 and it is evidently impossible to choose representatives naturally to get theorem 1.5, they are chosen merely to get the desired result. Another objection to this point of view was raised in the introduction, that of the burden of carrying an isomorphism through a problem.

Recently, much work has been done in generalizing the following theorem [7]:

If (Y,μ) is a Kakutani space and if $\mathcal{G}$ is a full σ-subfield of Baire sets, then there exist a Kakutani space (Z,ν), a continuous mapping π from Y onto Z, and, for each z in Z, a Baire measure μ^z in $\pi^{-1}(z)$ such that the set transformation induced by π is a one-one mapping from $\mathcal{G} \cap \mathcal{E}$ onto the class $\mathcal{F}$ of all open-closed sets in Z and such that, for every open-closed set E in Y

$$\mu(E) = \int \mu^z(E \cap \pi^{-1}(z))\, d\nu(z).$$

Most notable are the works of A. and C. Ionescu Tulcea [8]. In this area, the basic probability space, corresponding to Y, is a locally compact topological space and the central problem is the disintegration of probabilities. The probabilities μ^z defined above are actually representatives of conditional probabilities on the space Y. However, as was remarked in the introduction, a locally compact topological space must be rejected as the most general space to be considered for a probability space. Thus, the generalizations being made in the disintegration problem are not useful for conditional probability.

2. SOME ASPECTS OF PROBABILITY ON A BOOLEAN ALGEBRA

In this chapter only those topics of probability theory on a Boolean algebra which relate to conditioning are considered. This work was motivated by the lack of such a treatment in Kappos [12]. It should be remarked that a probability as defined in Kappos [12] is not used here and his probabilities will be called positive probability measures so that they may be distinguished from the probabilities introduced here.

2.1. Random Variables

A detailed exposition of the definitions and remarks of this section can be found in Olmsted [19] unless some other source is explicitly stated. In order to study conditional probability, it is advantageous to assume the Boolean algebra of events is σ-complete even though some of the theory can be carried out for more general classes of events (v. Varadarajan [24]).

Let B be a σ-complete Boolean algebra and let R denote the set of real numbers. Recall that $a \leq b$ means that $a \wedge b = a$, or equivalently $a \vee b = b$, where a and b are elements of B (v. Kappos [12]). A function $\hat{f}$ taking R into B is a random variable if the following conditions are satisfied:

1. $\hat{f}(\alpha) \downarrow$ as $\alpha \uparrow$,
2. $\bigvee_{\alpha} \hat{f}(\alpha) = 1$ and $\bigwedge_{\alpha} \hat{f}(\alpha) = 0$,
3. $\bigvee_{\beta > \alpha} \hat{f}(\beta) = \hat{f}(\alpha)$ for every α in R.

Note that if B is a σ-field of subsets of a space X, a one-one correspondence between random variables $\hat{f}$ and measurable real-valued functions f is given by $\hat{f}(\alpha) = \{x : f(x) > \alpha\}$. This will motivate the definitions that follow.

The order on random variables. $\hat{f} \leq \hat{g}$ means $\hat{f}(\alpha) \leq \hat{g}(\alpha)$ for every α in R, $\hat{f} = \hat{g}$ means $\hat{f}(\alpha) = \hat{g}(\alpha)$ for every α, and $\hat{f} < \hat{g}$ means $\hat{f} \leq \hat{g}$, but $\hat{f} \neq \hat{g}$. Under this order, the random variables form a partially ordered set. If $\hat{f}_n$, $\hat{g}$ are random variables, $\hat{f}_n \leq \hat{g}$, then $(\bigvee_n \hat{f}_n)(\alpha) = \bigvee_n \hat{f}_n(\alpha)$. If $\hat{f}_n \geq \hat{g}$, then $(\bigwedge_n \hat{f}_n)(\alpha) = \bigvee_{m=1}^{\infty} \bigwedge_n \hat{f}_n(\alpha + 1/m)$.

Addition of random variables. $\hat{f} + \hat{g}$ is defined to be the random variable defined by $(\hat{f} + \hat{g})(\alpha) = \bigvee_{\beta} \wedge[\hat{f}(\beta), \hat{g}(\alpha-\beta)]$, where β ranges over a countable dense set of real numbers. Then addition is commutative and associative.

Multiplication of random variables by real numbers. If c is a positive real number, $c\hat{f}$ is the random variable defined by $(c\hat{f})(\alpha) = \hat{f}(\alpha/c)$. $-\hat{f}$ is given by $(-\hat{f})(\alpha) = 1 - \bigwedge_{\beta < -\alpha} \hat{f}(\beta) = 1 - \bigwedge_{n=1}^{\infty} \hat{f}(-\alpha - 1/n)$. $\hat{0} = 0\hat{f}$ is given by

$$\hat{0}(\alpha) = \begin{cases} 1 , & \alpha < 0 \\ \\ 0 , & 0 \leq \alpha . \end{cases}$$

If $c < 0$, $c\hat{f}$ is the random variable $-[(-c)\hat{f}]$. Then $\alpha(\beta\hat{f}) = (\alpha\beta)\hat{f}$, $(\alpha+\beta)\hat{f} = \alpha\hat{f} + \beta\hat{f}$, $\alpha(\hat{f}+\hat{g}) = \alpha\hat{f} + \alpha\hat{g}$, for all real numbers α, β and for any random variables $\hat{f}, \hat{g}$. There is a unit random variable $\hat{1}$ given by

$$\hat{1}(\alpha) = \begin{cases} 1 \;, & \alpha < 1 \\ 0 \;, & 1 \leq \alpha \;. \end{cases}$$

With the above, the class of random variables form a σ-complete vector lattice (v. Birkhoff [2]).

Multiplication of random variables. If $\hat{f}, \hat{g} \geq \hat{0}$,

$$(\hat{f}\hat{g})(\alpha) = \begin{cases} 1 \;, & \alpha < 0 \\ \bigvee_{\beta > 0} \wedge[\hat{f}(\beta),\ \hat{g}(\alpha/\beta)], & 0 \leq \alpha, \end{cases}$$

where β ranges over a countable dense set of real numbers. If $\hat{f}^+ = \hat{f} \vee \hat{0}$ and $\hat{f}^- = -(\hat{f} \wedge \hat{0})$, define $\hat{f}\hat{g} = \hat{f}^+ \hat{g}^+ + \hat{f}^- \hat{g}^- - \hat{f}^+ \hat{g}^- - \hat{f}^- \hat{g}^+$. Then multiplication is commutative, associative, and multiplication distributes over addition. The usual properties of absolute value hold for $|\hat{f}| = \hat{f}^+ + \hat{f}^-$ (v. Birkhoff [2]).

Elements of the Boolean algebra correspond to random variables called characteristic functions and are defined by

$$\hat{a}(\alpha) = \begin{cases} 1 \;, & \alpha < 0 \\ a \;, & 0 \leq \alpha < 1 \\ 0 \;, & 1 \leq \alpha. \end{cases}$$

Note that the unit and zero random variables $\hat{1}$ and $\hat{0}$, respectively are two-valued characteristic functions.

A simple function is one which takes only a finite number of values. Any simple function can be represented as a linear combination of disjoint characteristic functions.

Limits of sequences of random variables. (v. Birkhoff [2]) $\hat{f}_n \to \hat{f}$ means $\limsup_n \hat{f}_n = \liminf_n \hat{f}_n = \hat{f}$, that is $\bigwedge_n\{\bigvee_{m\geq n} \hat{f}_m\} = \bigvee_n\{\bigwedge_{m\geq n} \hat{f}_m\} = \hat{f}$. The simple functions are dense in the class of random variables in the sense of this limit.

2.2. Probability and Integration

Let B be a σ-complete Boolean algebra and let $\Omega[0,1]$ be the set of random variables which assume only 0 and 1 of B. A mapping $\hat{\mu}$ taking B into $\Omega[0,1]$ is called a probability if

1. $\hat{0} \leq \hat{\mu}(a) \leq \hat{1}$ for every a in B,
2. $\hat{\mu}(a) = \hat{0}$ if and only if $a = 0$,
3. $\hat{\mu}(1) = \hat{1}$,
4. if $a \wedge b = 0$, $\hat{\mu}(a \vee b) = \hat{\mu}(a) + \hat{\mu}(b)$, and
5. if $a_n \downarrow 0$, $\hat{\mu}(a_n) \downarrow \hat{0}$

hold, where $\hat{\mu}(a)$ is the image of a in B under $\hat{\mu}$, that is $\hat{\mu}(a)$ is a random variable which assumes values 0 and 1. Condition 2 is a positivity condition and 4 and 5 are additivity and σ-additivity (continuity) conditions, respectively.

This notion of a probability is different from that of Kappos [12] and all other investigations of this topic, but they are closely related. If w is a positive probability measure on a σ-field of subsets of some space X, define $\hat{\mu}$ by

$$\hat{\mu}(A) = \begin{cases} X, & \alpha < w(A) \\ \phi, & w(A) \leq \alpha, \end{cases}$$

for each event A. Note that the real number at which the ''jump'' of the random variable $\hat{\mu}(A)$ occurs corresponds to the measure of A.

The notation $\hat{\Phi}$ will be used for a mapping of B into the class of all random variables $(\Omega[B])$ with the properties 1-5 above.

The next step will be to define integrals. They will be defined with respect to more general mappings than the analogues of measures. To be precise the integral will be defined relative to mappings described in the preceding paragraph. It should be pointed out that the integral will not be numerically valued, but the values are random variables.

If $\hat{s} = \alpha_1\hat{a}_1 + \ldots + \alpha_n\hat{a}_n$, $\alpha_i \geq 0$, $a_i \wedge a_j = 0$, $i \neq j$, $i,j=1,\ldots,n$, is a simple function, define its integral relative to the mapping $\hat{\Phi}$ by

$$\int \hat{s}\, d\, \hat{\Phi} = \alpha_1\, \hat{\Phi}(a_1) + \ldots + \alpha_n\, \hat{\Phi}(a_n).$$

Then the integral of a simple function is a linear combination of random variables.

If $\hat{f}$ is a non-negative random variable, that is $\hat{f} \geq \hat{0}$, define its integral by

$$\int \hat{f}\, d\, \hat{\Phi} = \sup\{\int \hat{s}\, d\, \hat{\Phi} : \hat{s} \text{ simple}, \ \hat{0} \leq \hat{s} \leq \hat{f}\}.$$

In order to prove some of the properties an integral should have, the following notation is useful. For a real number α, let $\hat{\alpha}$ denote the random variable given by

$$\hat{\alpha}(\beta) = \begin{cases} 1 & , \beta < \alpha \\ 0 & , \alpha \leq \beta . \end{cases}$$

Generally $\hat{\epsilon}$ and $\hat{N}$ will denote random variables defined above whose jump occurs at a small positive real number and a large real number, respectively.

The theorems 2.2, 2.3 and 2.4 are generalizations of theorems in Olmsted [19], where complete proofs are not given.

Lemma2.1. If $a < b$, $\hat{\Phi}(a) \leq \hat{\Phi}(b)$.

Proof. If $a \leq b$, $b = a \vee (b-a)$ so by 4 of the definition of $\hat{\Phi}$, $\hat{\Phi}(b) = \hat{\Phi}(a) + \hat{\Phi}(b-a)$, and then by 1 of the definition of $\hat{\Phi}$, $\hat{\Phi}(b) \geq \hat{\Phi}(a)$.

Theorem 2.2. If $\hat{0} \leq \hat{f} \leq \hat{g}$, $\int \hat{f}\, d\, \hat{\Phi} \leq \int \hat{g}\, d\, \hat{\Phi}$.

Proof. At first suppose that $\hat{f}$ and $\hat{g}$ are characteristic functions, then $\hat{f} = \hat{a}$ for some a in B and $\hat{g} = \hat{b}$ for some b in B such that $a \leq b$. Then $\int \hat{f}\, d\, \hat{\Phi} = \hat{\Phi}(a)$ and $\int \hat{g}\, d\, \hat{\Phi} = \hat{\Phi}(b)$. By lemma 2.1, $\int \hat{f}\, d\, \hat{\Phi} \leq \int \hat{g}\, d\, \hat{\Phi}$.

If $\hat{f}$ and $\hat{g}$ are simple functions, there is a common decomposition so that

$$\hat{f} = \alpha_1 \hat{a}_1 + \ldots + \alpha_n \hat{a}_n \quad \text{and} \quad \hat{g} = \beta_1 \hat{a}_1 + \ldots + \beta_n \hat{a}_n,$$

where $\beta_i \geq \alpha_i$, $i=1,\ldots,n$ (v. Olmsted [19]). Now it is clear that $\int \hat{f}\, d\, \hat{\Phi} \leq \int \hat{g}\, d\, \hat{\Phi}$ in this case.

For non-negative random variables the result is obvious in view of the definition of the integral.

<u>Theorem 2.3</u>. If $\hat{s}$ and $\hat{t}$ are simple non-negative functions and α is a positive real number, $\int(\hat{s}+\hat{t})d\hat{\Phi} = \int\hat{s}\, d\, \hat{\Phi} + \int\hat{t}\, d\, \hat{\Phi}$, and $\int\alpha\, \hat{s}\, d\, \hat{\Phi} = \alpha\int\hat{s}\, d\, \hat{\Phi}$.

<u>Proof</u>. Let $\hat{s} = \alpha_1\hat{a}_1 +\ldots+ \alpha_n\hat{a}_n$, then $\alpha\hat{s} = \alpha\alpha_1\hat{a}_1+\ldots+\alpha\alpha_n\hat{a}_n$ and $\int\alpha\, \hat{s}\, d\,\hat{\Phi} = \alpha\alpha_1\hat{\Phi}(a_1) +\ldots+ \alpha\alpha_n\hat{\Phi}(a_n) = \alpha\int\hat{s}\, d\, \hat{\Phi}$.

There is a common decomposition so that $\hat{s} = \alpha_1\hat{a}_1+\ldots+\alpha_n\hat{a}_n$ and $\hat{t} = \beta_1\hat{a}_1+\ldots+ \beta_n\hat{a}_n$, and $\hat{s} + \hat{t} = (\alpha_1+\beta_1)\hat{a}_1+\ldots+(\alpha_n+\beta_n)\hat{a}_n$ (v. Olmsted [19]). Then $\int(\hat{s}+\hat{t})d\,\hat{\Phi} = (\alpha_1+\beta_1)\,\hat{\Phi}(a_1) +\ldots+(\alpha_n + \beta_n)\,\hat{\Phi}(a_n)$ and by the properties of multiplication by real numbers

$$\int(\hat{s}+\hat{t})\, d\,\hat{\Phi} = \alpha_1\hat{\Phi}(a_1) +\ldots+ \alpha_n\hat{\Phi}(a_n) + \beta_1\hat{\Phi}(a_1) +\ldots+ \beta_n\hat{\Phi}(a_n)=\int\hat{s}\, d\,\hat{\Phi} + \int\hat{t}\, d\hat{\Phi}.$$

<u>Theorem 2.4</u>. If $\hat{s}_n$ is an increasing sequence of non-negative simple functions whose limit is the random variable $\hat{f}$, then $\int\hat{s}_n d\,\hat{\Phi} \uparrow \int\hat{f}\, d\,\hat{\Phi}$.

<u>Proof</u>. Let $\hat{s}$ be a simple function for which $\hat{0} \leq \hat{s} \leq \hat{f}$ and if $\hat{s}_n$ is as in the statement of the theorem, then $\lim\limits_n \hat{s}_n \geq \hat{s}$. Since $(\bigvee\limits_n \wedge[\hat{s},\hat{s}_n])(\alpha) = \bigvee\limits_n(\wedge[\hat{s},\hat{s}_n])(\alpha) = \bigvee\limits_n(\hat{s}(\alpha) \wedge \hat{s}_n(\alpha)) = \hat{s}(\alpha)$, $\wedge[\hat{s},\hat{s}_n] \uparrow \hat{s}$. Then

$$(\hat{s} - \wedge[\hat{s},\, \hat{s}_n]) = \hat{t}_n \downarrow \hat{0}.$$

The $\hat{t}_n$ are simple functions so if $N = \sup\{\alpha:\hat{t}_1(\alpha) > 0\}$, $\hat{t}_n\leq \hat{N}$ for every n. Let $\epsilon > 0$, then

$$\hat{t}_n \leq \hat{\epsilon} + N(\widehat{\hat{t}_n(\epsilon)}).$$

By theorems 2.2 and 2.3

$$\int \hat{t}_n d\ \hat{\Phi} \leq \int \hat{\epsilon}\ d\ \hat{\Phi} + \int N(\widehat{\hat{t}_n(\epsilon)}) d\ \hat{\Phi} = \hat{\epsilon} + N\ \hat{\Phi}(\hat{t}_n(\epsilon)).$$

But $\hat{t}_n(\epsilon) \downarrow 0$ so by 5 of the requirements on $\hat{\Phi}$,

$$\hat{\Phi}(\hat{t}_n(\epsilon)) \downarrow \hat{0}.$$

Thus, letting $\epsilon \downarrow 0$,

$$\lim_n \int (\hat{s} - \wedge[\hat{s}, \hat{s}_n]) d\ \hat{\Phi} = \hat{0}.$$

By theorem 2.2 and 2.3

$$\int \hat{s}\ d\ \hat{\Phi} = \lim_n \int \wedge[\hat{s}, \hat{s}_n] d\ \hat{\Phi} \leq \lim_n \int \hat{s}_n d\ \hat{\Phi}.$$

Then

$$\int \hat{s}\ d\ \hat{\Phi} \leq \lim_n \int \hat{s}_n d\ \hat{\Phi} \leq \int \hat{f}\ d\ \hat{\Phi}$$

and taking the supremum over the $\hat{s}$ for which $\hat{0} \leq \hat{s} \leq \hat{f}$, the desired result is obtained.

A random variable $\hat{f}$ is bounded above if there is α such that $\hat{f}(\alpha) = 0$; $\hat{f}$ is bounded below if there is β such that $\hat{f}(\beta) = 1$; $\hat{f}$ is bounded if it is bounded above and bounded below.

The definition of the integral for arbitrary random variables and the following theorems 2.5 and 2.6 do not differ from the standard ones, but will be included for the sake of completeness.

<u>Theorem 2.5</u>. If $\hat{f}$ and $\hat{g}$ are non-negative random variables whose integrals are bounded and α is a positive real number, then $\int \alpha \hat{f} \, d \hat{\Phi}$, and $\int (\hat{f}+\hat{g}) d \hat{\Phi}$ are bounded, and $\int \alpha \hat{f} \, d \hat{\Phi} = \alpha \int \hat{f} \, d \hat{\Phi}$, $\int (\hat{f}+\hat{g}) d \hat{\Phi} = \int \hat{f} \, d \hat{\Phi} + \int \hat{g} \, d \hat{\Phi}$.

<u>Proof</u>. Let $\hat{s}_n$ and $\hat{t}_n$ be sequences of non-negative simple functions such that $\hat{s}_n \uparrow \hat{f}$ and $\hat{t}_n \uparrow \hat{g}$. Then $\alpha\hat{s}_n \uparrow \alpha\hat{f}$ (v. Olmsted [19]) and by theorems 2.3 and 2.4

$$\int \alpha \hat{f} \, d \hat{\Phi} = \lim_n \int \alpha \hat{s}_n d \hat{\Phi} = \alpha \int \hat{f} \, d \hat{\Phi}.$$

Since $\hat{0} \leq \hat{s}_n + \hat{t}_n \uparrow \hat{f} + \hat{g}$ (v. Olmsted [19])

$$\lim_n \int (\hat{s}_n + \hat{t}_n) \, d \hat{\Phi} = \int (\hat{f} + \hat{g}) d \hat{\Phi}$$

making use of theorem 2.4. But by theorems 2.2 and 2.3,

$$\int (\hat{s}_n + \hat{t}_n) d \hat{\Phi} = \int \hat{s}_n d \hat{\Phi} + \int \hat{t}_n d \hat{\Phi} \leq \int \hat{f} \, d \hat{\Phi} + \int \hat{g} \, d \hat{\Phi} .$$

On the other hand,

$$\int (\hat{f} + \hat{g}) d \hat{\Phi} \geq \int (\hat{s}_n + \hat{t}_n) d \hat{\Phi} = \int \hat{s}_n d \hat{\Phi} + \int \hat{t}_n d \hat{\Phi}.$$

But (v. Birkhoff [2])

$$\lim_n \left(\int \hat{s}_n d \hat{\Phi} + \int \hat{t}_n d \hat{\Phi}\right) = \int \hat{f} \, d \hat{\Phi} + \int \hat{g} \, d \hat{\Phi}.$$

Combining the two implied inequalities, the proof is terminated.

A random variable $\hat{f}$ is integrable when the integrals of $\hat{f}^+$ and $\hat{f}^-$ are bounded and its integral is defined as

$$\int \hat{f}\, d\,\hat{\Phi} = \int \hat{f}^+\, d\,\hat{\Phi} - \int \hat{f}^-\, d\,\hat{\Phi}\,.$$

Theorem 2.6. If $\hat{f}$ and $\hat{g}$ are integrable, then $\alpha\,\hat{f}$ and $\hat{f} + \hat{g}$ are integrable and

$$\int \alpha\,\hat{f}\, d\,\hat{\Phi} = \alpha \int \hat{f}\, d\,\hat{\Phi}, \quad \text{and}$$

$$\int (\hat{f} + \hat{g})\, d\,\hat{\Phi} = \int \hat{f}\, d\,\hat{\Phi} + \int \hat{g}\, d\,\hat{\Phi}\,.$$

Proof. If $\hat{f}$ and $\hat{g}$ are integrable, it follows from the definition of integrable that $|\hat{f}|$ and $|\hat{g}|$ are integrable. But

$$(\hat{f} + \hat{g})^+ \leq |\hat{f}| + |\hat{g}| \quad \text{and} \quad (\hat{f}+\hat{g})^- \leq |\hat{f}| + |\hat{g}|$$

so by theorem 2.2 and 2.5 $\hat{f} + \hat{g}$ is integrable. By an application of theorem 2.5 to (v. Olmsted [19])

$$(\hat{f} + \hat{g})^+ + \hat{f}^- + \hat{g}^- = (\hat{f} + \hat{g})^- + \hat{f}^+ + \hat{g}^+,$$

$$\int (\hat{f}+\hat{g})^+\, d\,\hat{\Phi} - \int (\hat{f}+\hat{g})^-\, d\,\hat{\Phi} = \int \hat{f}^+ d\,\hat{\Phi} - \int \hat{f}^- d\,\hat{\Phi} + \int \hat{g}^+ d\,\hat{\Phi} - \int \hat{g}^- d\,\hat{\Phi}.$$

Since $\alpha\hat{f} = \alpha\hat{f}^+ - \alpha\hat{f}^-$, by means of the above and theorem 2.5,

$$\int \alpha\hat{f}\, d\hat{\Phi} = \alpha \int \hat{f}^+\, d\hat{\Phi} - \alpha \int \hat{f}^-\, d\hat{\Phi} = \alpha \int \hat{f}\, d\,\hat{\Phi}\,.$$

The theorems 2.7 - 2.9 are proved exactly as in standard measure theory (for example, v. Loeve [16]) now that the preliminary theorems have been established.

Theorem 2.7. (Monotone Convergence Theorem). If $\hat{0} \leq \hat{f}_n$ is integrable and $\hat{f}_n \uparrow \hat{f}$, then $\hat{f}$ is integrable and

$$\int \hat{f}_n \, d\, \hat{\Phi} \uparrow \int \hat{f} \, d\, \hat{\Phi} .$$

Theorem 2.8. (Fatou-Lebesgue Theorem). Let $\hat{h}$ and $\hat{g}$ be integrable. If $\hat{h} \leq \hat{f}_n$ or $\hat{f}_n \leq \hat{g}$, then

$\int \liminf_n \hat{f}_n d\, \hat{\Phi} \leq \liminf_n \int \hat{f}_n \, d\, \hat{\Phi}$, respectively

$\limsup_n \int \hat{f}_n d\, \hat{\Phi} \leq \int \limsup_n \hat{f}_n \, d\, \hat{\Phi}$.

Theorem 2.9. (Dominated Convergence Theorem). If $|\hat{f}_n| \leq \hat{g}$ with $\hat{g}$ integrable and if $\hat{f}_n \to \hat{f}$, then

$$\int \hat{f}_n d\, \hat{\Phi} \to \int \hat{f}_n \, d\, \hat{\Phi} .$$

2.3. Absolute Continuity

Let B be a σ-complete Boolean algebra and let $\hat{\mu}$ be a probability. A mapping $\hat{\eta}$ taking B into $\Omega[0,1]$ (v. section 2.2) is absolutely continuous with respect to $\hat{\mu}$ in case, given any $\hat{\epsilon} > \hat{0}$, there is $\hat{\delta} > \hat{0}$ such that

$$|\hat{\eta}(a)| < \hat{\epsilon} \quad \text{whenever} \quad \hat{\mu}(a) < \hat{\delta},$$

where a is an element of B.

The integral of an integrable random variable $\hat{f}$ over an element a of the Boolean algebra B is defined by

$$\int_a \hat{f}\, d\,\hat{\Phi} = \int \hat{a}\,\hat{f}\, d\,\hat{\Phi},$$

where $\hat{\Phi}$ is a mapping as in the previous section.

<u>Theorem 2.10</u>. If $\hat{f}$ is an integrable random variable, the following hold:

1. if $\hat{f} \geq \hat{0}$, $\int_a \hat{f}\, d\,\hat{\Phi} \geq \hat{0}$ for every a,
2. if $\hat{f} > \hat{0}$, $\int_a \hat{f}\, d\,\hat{\Phi} > \hat{0}$ for some a,
3. if $\hat{f}$ not $\geq \hat{0}$, $\int_a \hat{f}\, d\,\hat{\Phi} < \hat{0}$ for some a, and
4. if $\int_a \hat{f}\, d\,\hat{\Phi} = \hat{0}$ for every a, $\hat{f} = \hat{0}$.

<u>Proof</u>. (1) If $\hat{f} \geq \hat{0}$, $\hat{0} \leq \hat{a}\hat{f} \leq \hat{f}$ for every a. Let $\alpha = \sup\{\beta : (\hat{a}\hat{f})(\beta) = a\}$, then $\hat{0} \leq \alpha\,\hat{a} \leq \hat{a}\,\hat{f}$, that is

$$\hat{0} \leq \alpha\,\hat{\Phi}\,(a) \leq \int_a \hat{f}\, d\,\hat{\Phi}\,.$$

(2) If $\hat{f} > \hat{0}$, there is $\alpha > 0$ for which $a = \hat{f}(\alpha) > 0$. Then $0 < \alpha\,\hat{a} \leq \hat{a}\,\hat{f}$, and $\hat{0} < \alpha\,\hat{\Phi}(a) \leq \int_a \hat{f}\, d\,\hat{\Phi}$.

(3) If $\hat{f}$ not $\geq \hat{0}$, there is $\alpha < 0$ for which $a = \hat{f}(\alpha) < 1$. Then $\hat{a}\,\hat{f} \leq \alpha\,\hat{a}$, and $\int_a \hat{f}\, d\,\hat{\Phi} \leq \alpha\,\hat{\Phi}\,(a) < \hat{0}$.

(4) If $\hat{f} \neq \hat{0}$, $\hat{f}$ not $\geq \hat{0}$ or $\hat{f}$ not $\leq \hat{0}$. In either case an application of 3 leads to a contradiction and the desired result is attained.

Note that a probability $\hat{\mu}$ is a less general mapping than $\hat{\Phi}$. Thus the general integration theory is already established. We turn now to the Radon-Nikodym theorem. Although the integral of a random variable with respect to a probability is not the same as in Olmsted [19], the proof of the Radon-Nikodym can be done in the same manner. Since it is primarily a matter of substituting $\hat{\alpha}$ for α and mappings from B into $\Omega[0,1]$ for mappings from B into the real numbers, the proof will not be given here. The Radon-Nikodym theorem is stated as follows.

> Theorem 2.11. If $\hat{\eta}$ is absolutely continuous with respect to $\hat{\mu}$, there is a unique random variable $\hat{f}$ such that $\hat{\eta}(a) = \int_a \hat{f}\, d\,\hat{\mu}$. Conversely for a random variable $\hat{f}$ the integral $\hat{\eta}(a) = \int_a \hat{f}\, d\,\hat{\mu}$ is absolutely continuous.

The Radon-Nikodym theorem lays the groundwork for conditioning in probability theory and this topic will be considered in the next section.

2.4. Conditioning

Let B_o be a Boolean σ-subalgebra of B. The restriction of $\hat{\mu}$ to B_o, written $\hat{\mu}|_{B_o}$, is a probability on B_o. If $\hat{f}$ is an integrable random variable, the mapping $\hat{\eta}$ defined by

$$\hat{\eta}(a) = \int_a \hat{f}\, d\,\hat{\mu}$$

for a in B_o is absolutely continuous with respect to $\hat{\mu}$ by theorem 2.11. Therefore $\hat{\eta}$ is absolutely continuous with respect to $\hat{\mu}|_{B_o}$ so there is a random variable $\hat{E}_{B_o}(\hat{f})$ taking values in B_o for which

$$\int_a \hat{E}_{B_o}(\hat{f})\, d\hat{\mu}|_{B_o} = \hat{\eta}(a)$$

for every a in B_o. Let $\hat{E}_{B_o}$ denote the mapping which takes a random variable $\hat{f}$ into $\hat{E}_{B_o}(\hat{f})$. $\hat{E}_{B_o}$ will be called a conditional expectation given B_o and $\hat{E}_{B_o}(\hat{f})$ is the conditional expectation of $\hat{f}$, given B_o.

<u>Theorem 2.12</u>. $\hat{E}_{B_o}$ satisfies the following conditions:

C.E.1. $\hat{E}_{B_o}(\hat{1}) = \hat{1}$,

C.E.2. if $\hat{f} \geq \hat{0}$, $\hat{E}_{B_o}(\hat{f}) \geq \hat{0}$,

C.E.3. $\hat{E}_{B_o}(\sum_{j=1}^{n} c_j \hat{f}_j) = \sum_{j=1}^{n} c_j \hat{E}_{B_o}(\hat{f}_j)$,

C.E.4. $|\hat{E}_{B_o}(\hat{f})| \leq \hat{E}_{B_o}(|\hat{f}|)$, and

C.E.5. if $\hat{f}_n \to \hat{f}$ and there is $\hat{g}$ integrable such that $|\hat{f}| \leq \hat{g}$, then $\hat{E}_{B_o}(\hat{f}_n) \to \hat{E}_{B_o}(\hat{f})$.

<u>Proof</u>. (C.E.1) $\hat{1}$ takes values in B_o so $\hat{E}_{B_o}(\hat{1}) = \hat{1}$.

(C.E.2) If $\hat{f} \geq \hat{0}$, $\int_a \hat{f}\, d\hat{\mu} \geq \hat{0}$ for every a by 1 of theorem 2.10.

Then

$$\int_a \hat{E}_{B_o}(\hat{f})\, d\hat{\mu}|_{B_o} \geq \hat{0}$$

for every a in B_o. If $\hat{E}_{B_o}(\hat{f})$ not $\geq \hat{0}$, there is a in B_o such that $\int_a \hat{E}_{B_o}(\hat{f})\, d\,\hat{\mu}|_{B_o} < \hat{0}$ by 3 of theorem 2.10. Thus $\hat{E}_{B_o}(\hat{f}) \geq 0$.

(C.E.3) Note that

$$\int_a \sum_{j=1}^{n} c_j \hat{E}_{B_o}(\hat{f}_j)\, d\,\hat{\mu}|_{B_o} = \sum_{j=1}^{n} c_j \int_a \hat{E}_{B_o}(\hat{f}_j)\, d\,\hat{\mu}|_{B_o}$$

$$= \sum_{j=1}^{n} c_j \int_a \hat{f}_j\, d\,\hat{\mu} = \int_a \sum_{j=1}^{n} c_j \hat{f}_j\, d\,\hat{\mu} \quad \text{for every } a \text{ in } B_o.$$

But then

$$\hat{E}_{B_o}\Big(\sum_{j=1}^{n} c_j \hat{f}_j\Big) = \sum_{j=1}^{n} c_j \hat{E}_{B_o}(\hat{f}_j).$$

(C.E.4) Since $|\hat{f}| + \hat{f}$ and $|\hat{f}| - \hat{f}$ are $\geq \hat{0}$, $\hat{E}_{B_o}(|\hat{f}| + \hat{f})$ and $\hat{E}_{B_o}(|\hat{f}| - \hat{f})$ are $\geq \hat{0}$ by C.E.2. Then by C.E. 3

$$\hat{E}_{B_o}(|\hat{f}|) \geq \hat{E}_{B_o}(\hat{f}) \quad \text{and} \quad \hat{E}_{B_o}(|\hat{f}|) \geq -\hat{E}_{B_o}(\hat{f}).$$

Therefore $\hat{E}_{B_o}(|\hat{f}|) \geq |\hat{E}_{B_o}(\hat{f})|$.

(C.E.5) If $\hat{f}_n \to \hat{f}$, $\hat{y}_n = \bigvee_{j \geq n} |\hat{f}_j - \hat{f}|$ converges downward to $\hat{0}$ (v. Birkhoff [2]). Moreover, $\hat{y}_n \leq 2\,\hat{g}$ and

$$|\hat{E}_{B_o}(\hat{f}_n) - \hat{E}_{B_o}(\hat{f})| = |\hat{E}_{B_o}(\hat{f}_n - \hat{f})| \leq \hat{E}_{B_o}(|\hat{f}_n - \hat{f}|) \leq \hat{E}_{B_o}(\hat{y}_n).$$

Now $\hat{E}_{B_o}(\hat{y}_1) \geq \hat{E}_{B_o}(\hat{y}_2) \geq \dots \geq \hat{0}$ and $\hat{E}_{B_o}(\hat{y}_n)$ converges to some $\hat{y} \geq \hat{0}$.

But

$$\int \hat{y}\, d\,\hat{\mu}|_{B_o} \leq \int \hat{E}_{B_o}(\hat{y}_n)\, d\,\hat{\mu}|_{B_o} = \int \hat{y}_n\, d\,\hat{\mu},$$

so by the Fatou-Lebesgue theorem

$$\hat{0} \leq \int \hat{y}\, d\hat{\mu}|_{B_o} \leq \lim_n \int \hat{y}_n\, d\hat{\mu} = \hat{0}.$$

Thus $\hat{y} = \hat{0}$ by 4 of theorem 2.10 and (v. Birkhoff [2])

$$\hat{E}_{B_o}(\hat{f}_n) \to \hat{E}_{B_o}(\hat{f}).$$

This concludes the proof of theorem 2.12.

The properties C.E.1-C.E.5 of conditional expectation resemble those of the traditional conditional expectation as given by Doob [5]. However, there is an important difference other than the fact that the functions and integrals are different, namely there is no qualifying ''with probability one'' appended to the properties as in Doob. The importance of this difference is brought out in the properties of conditional probability, to be defined, and theorem 2.14.

If $\hat{f}$ is the characteristic function of an element of B, that is $\hat{f} = \hat{a}$, $\hat{E}_{B_o}(\hat{a})$ will be written $\hat{\mu}_{B_o}(a)$ and called the conditional probability of a, given B_o. The mapping $\hat{E}_{B_o}$ restricted to the characteristic functions of B is written $\hat{\mu}_{B_o}$ and called a conditional probability given B_o. Note that the defining equation becomes

$$\int_a \hat{\mu}_{B_o}(b)\, d\hat{\mu}|_{B_o} = \int_a \hat{b}\, d\,\hat{\mu} = \hat{\mu}(a \wedge b),$$

for every a in B_o.

Theorem 2.13. $\hat{\mu}_{B_o}$ takes its values in $\Omega[B_o]$ and satisfies:

C.P.1. $\hat{0} \leq \hat{\mu}_{B_o}(a) \leq \hat{1}$ for every a in B,

C.P.2. $\hat{\mu}_{B_o}(a) = \hat{0}$ if and only if $a = 0$,

C.P.3. $\hat{\mu}_{B_o}(1) = \hat{1}$,

C.P.4. if $a \wedge b = 0$, $\hat{\mu}_{B_o}(a \vee b) = \hat{\mu}_{B_o}(a) + \hat{\mu}_{B_o}(b)$, and

C.P.5. if $a_n \downarrow 0$, $\hat{\mu}_{B_o}(a_n) \downarrow \hat{0}$.

Proof. $\hat{\mu}_{B_o}(1) = \hat{E}_{B_o}(\hat{1}) = \hat{1}$ by C.E.1 so C.P.3 is obtained.

By definition

$$\int_a \hat{\mu}_{B_o}(0)\, d\, \hat{\mu}|_{B_o} = \hat{\mu}(a \wedge 0) = \hat{0}$$

for every a in B_o, so by 4 of theorem 2.10, $\hat{\mu}_{B_o}(0) = \hat{0}$. If $\hat{\mu}_{B_o}(b) = \hat{0}$, $\int_a \hat{\mu}_{B_o}(b)\, d\, \hat{\mu}|_{B_o} = \hat{\mu}(a \wedge b) = \hat{0}$, for every a in B_o, that is $b = 0$. Thus C.P.2 is obtained. For every a in B_o, $\hat{0} \leq \hat{a}$ so by C.E.2

$$\hat{0} = \hat{\mu}_{B_o}(0) = \hat{E}_{B_o}(\hat{0}) \leq \hat{E}_{B_o}(\hat{a}) = \hat{\mu}_{B_o}(a).$$

Using C.E.3

$$\hat{0} \leq \hat{\mu}_{B_o}(1-a) = \hat{E}_{B_o}(\hat{1}) - \hat{E}_{B_o}(\hat{a}) = \hat{1}-\hat{\mu}_{B_o}(a),$$

that is $\hat{\mu}_{B_o}(a) \leq \hat{1}$ for every a in B_o and C.P.1 is obtained. If $a \wedge b = 0$, $(\hat{a} + \hat{b}) = \widehat{(a \vee b)}$. Then using C.E.3

$$\hat{\mu}_{B_o}(a \vee b) = \hat{E}_{B_o}(\hat{a} + \hat{b}) = \hat{E}_{B_o}(\hat{a}) + \hat{E}_{B_o}(\hat{b}) = \hat{\mu}_{B_o}(a) + \hat{\mu}_{B_o}(b).$$

Only C.P.5 remains to be proved. If $a_n \downarrow 0$, $\hat{a}_n \downarrow \hat{0}$ so as is seen in the proof of C.E.5

$$\hat{\mu}_{B_o}(a_n) = \hat{E}_{B_o}(\hat{a}_n) \downarrow \hat{E}_{B_o}(\hat{0}) = \hat{\mu}_{B_o}(0) = \hat{0}.$$

The proof of theorem 2.13 is completed.

The reader will note that the properties C.P.1 - C.P.5 of the mapping $\hat{\mu}_{B_o}$ are exactly those of the requirements on a mapping with respect to which integrals were defined in section 2.2. In fact, the conditional probability is the motivation for defining the integrals as they were defined there. What other applications such integrals have is not known and will not be discussed here.

Note that there is no condition of regularity for conditional probability in this sense. Thus the existence of regular conditional probabilities and related problems does not plague the theory. Following is an important theorem of probability which cannot be obtained in general in the conventional theory because regularity is not generally the case (cf. Doob [5]).

<u>Theorem 2.14</u>. If $\hat{f}$ is an integrable random variable, $\hat{E}_{B_o}(\hat{f}) = \int \hat{f} \, d\, \hat{\mu}_{B_o}$.

<u>Proof.</u> If $\hat{f}$ is a non-negative simple function and

$$\hat{f} = \alpha_1 \hat{a}_1 + \ldots + \alpha_n \hat{a}_n,$$

then by C.E.3

$$\hat{E}_{B_o}(\hat{f}) = \alpha_1 \hat{E}_{B_o}(\hat{a}_1) + \ldots + \alpha_n \hat{E}_{B_o}(\hat{a}_n) = \alpha_1 \hat{\mu}_{B_o}(a_1) + \ldots + \alpha_n \hat{\mu}_{B_o}(a_n)$$

$$= \int \hat{f} \, d\, \hat{\mu}_{B_o}.$$

If $\hat{f}$ is non-negative, let $\hat{0} \leq \hat{s}_n \uparrow \hat{f}$, then

$$\int \hat{s}_n \, d\, \hat{\mu}_{B_o} \uparrow \int \hat{f} \, d\, \hat{\mu}_{B_o}, \text{ and } \hat{E}_{B_o}(\hat{s}_n) \uparrow \hat{E}_{B_o}(\hat{f}).$$

But $\int \hat{s}_n \, d\, \hat{\mu}_{B_o} = \hat{E}_{B_o}(\hat{s}_n)$ so $\int \hat{f} \, d\, \hat{\mu}_{B_o} = \hat{E}_{B_o}(\hat{f})$. If $\hat{f}$ is an arbitrary integrable function,

$$\int \hat{f} \, d\, \hat{\mu}_{B_o} = \int \hat{f}^+ \, d\, \hat{\mu}_{B_o} - \int \hat{f}^- d\, \hat{\mu}_{B_o} = \hat{E}_{B_o}(\hat{f}^+) - \hat{E}_{B_o}(\hat{f}^-)$$

$$= \hat{E}_{B_o}(\hat{f}^+ - \hat{f}^-) = \hat{E}_{B_o}(\hat{f}).$$

Theorem 2.14 is important in applications which the student of elementary probability views with no surprise, but it is not true in the conventional theory.

After the preliminaries, the concept of conditional probability is much more simple than in the conventional set-up. There is an apparent inability to correspond numerical values to events, random variables, etc. This is only apparent and the problem is considered in the next chapter. Note that none of the objections to the Stone space approach apply here.

3. RECONCILIATION WITH THE CONVENTIONAL THEORY

3.1. The Relation of Probability Measures on a σ-field of Sets to Probabilities on a σ-complete Boolean Algebra

If $(X, \mathcal{A}, \lambda)$ is a probability space in the sense of Kolmogorov, then $\mathcal{A}/\lambda$ ($\mathcal{A}$ modulo the σ-ideal of λ-null sets in $\mathcal{A}$) is a complete Boolean algebra whose elements are denoted by $[A]$, indicating A is an element of the residue class $[A]$. Thus 1 will be denoted by $[X]$ and 0 by $[\phi]$. A mapping in the sense of section 2.2 of chapter two can be defined on $\mathcal{A}/\lambda$ as follows:

$$(\hat{\mu}([A]))(\alpha) = \begin{cases} [X] \quad , & \alpha < \lambda(A) \\ \\ [\phi] \quad , & \lambda(A) \leq \alpha \ . \end{cases}$$

The following theorems 3.1 and 3.2 are obvious after a moments reflection, but proofs are given to aid the intuition.

<u>Theorem 3.1</u>. If $\hat{\mu}$ is defined as above, $\hat{\mu}$ is a probability on $\mathcal{A}/\lambda$ in the sense of section 2.2.

<u>Proof</u>. Just the positivity, additivity and continuity conditions will be checked.

If $\hat{\mu}([A]) = \hat{0}$, it is clear from the definition that $\lambda(A) = 0$. Then $[A] = [\phi]$.

If $[A] \wedge [B] = [\phi]$,

$$(\hat{\mu}([A]) + \hat{\mu}([B]))(\alpha) = \bigvee_{\beta} \wedge [(\hat{\mu}([A]))(\beta),\ (\hat{\mu}([B]))(\alpha-\beta)]$$

$$= \bigvee \begin{cases} \bigvee_{\alpha-\lambda(B)<\beta} \wedge[(\hat{\mu}([A]))(\beta),\ [X]] \\ \bigvee_{\beta \leq \alpha-\lambda(B)} \wedge[(\hat{\mu}([A]))(\beta),\ [\phi]] \end{cases}$$

$$= \bigvee_{\alpha-\lambda(B)<\beta} (\hat{\mu}([A]))(\beta) = (\hat{\mu}([A]))(\alpha-\lambda(B))$$

$$= \begin{cases} [X] \ , & \alpha-\lambda(B) < \lambda(A) \\ [\phi] \ , & \alpha-\lambda(B) \geq \lambda(A) \end{cases} = (\hat{\mu}([A] \vee [B]))(\alpha).$$

If $[A_n] \downarrow [\phi]$, for any $\alpha > 0$ there is n_o such that $n_o < n$ implies $\lambda(A_n) < \alpha$. Then

$$(\hat{\mu}([A_n]))(\alpha) = [\phi] \quad \text{for} \quad n > n_o,$$

that is

$$\bigvee_{m=1}^{\infty} [\bigwedge_n (\hat{\mu}([A_n]))(\beta+1/m)] = [\phi] \quad \text{for} \quad \beta \geq 0$$

so that

$$(\bigwedge_n \hat{\mu}([A_n]))(\beta) = [\phi] \quad \text{for} \quad \beta \geq 0.$$

If $\beta < 0$, $\beta + 1/m_o < 0$ for some m_o and

$$(\hat{\mu}([A_n]))(\beta+1/m_o) = [X] \quad \text{for all } n.$$

Then

$$(\bigwedge_n \hat{\mu}([A_n]))(\beta) = [X],$$

hence $\bigwedge_n \hat{\mu}([A_n]) = \hat{0}$ if $[A_n] \downarrow [\phi]$.

<u>Theorem 3.2</u>. If $\hat{\mu}$ is a probability on $\mathfrak{A}/\lambda$ in the sense of section 2.2, then there is a probability measure on the field of sets $\mathfrak{A}$ which corresponds to it as in theorem 3.1.

<u>Proof</u>. For A in $[A]$, define

$$w(A) = \sup\{\alpha : (\hat{\mu}([A]))(\alpha) = [X]\}.$$

Then it is clear from the requirements 1-3 on $\hat{\mu}$ that (1) $0 \leq w(A) \leq 1$ for every A in $\mathfrak{A}$, (2) $w(\phi) = 0$, and (3) $w(X) = 1$. (4) If $A_1 \wedge A_2 = \phi$, $[A_1] \wedge [A_2] = [\phi]$ and

$$w(A_1 \cup A_2) = \sup\{\alpha : \bigvee_\beta \wedge[(\hat{\mu}([A_1]))(\beta), (\hat{\mu}([A_2]))(\alpha-\beta)] = [X]\}.$$

Then

$$w(A_1 \cup A_2) = \sup\left\{\alpha : \bigvee \left\{\begin{array}{l} \bigvee_{\alpha - w(A_2) < \beta} \wedge[(\hat{\mu}([A_1]))(\beta),\ [X]] \\ \\ \bigvee_{\alpha - w(A_2) > \beta} \wedge[(\hat{\mu}([A_2]))(\beta),\ [\phi]] \end{array}\right\} = [X]\right\}$$

$$= \sup\{\alpha : (\hat{\mu}([A_1]))(\alpha - w(A_2)) = [X]\} = w(A_2) + w(A_1).$$

(5) If $A_n \downarrow \phi$, $[A_n] \downarrow [\phi]$ so $\hat{\mu}([A_n]) \downarrow \hat{0}$. For $\epsilon > 0$,

$$(\bigwedge_n \hat{\mu}([A_n]))(\epsilon/2) = \bigvee_{m=1}^{\infty} \bigwedge_n (\hat{\mu}([A_n])(\epsilon/2 + 1/m) = [\phi]$$

so for every m, $(\hat{\mu}([A_n]))(\epsilon/2+1/m) = [\phi]$ from some n on because $\hat{\mu}([A_n]) \geq \hat{\mu}([A_{n+1}])$. Thus there is n_o such that $(\hat{\mu}([A_n]))(\epsilon) = [\phi]$, that is $w(A_n) < \epsilon$ for $n > n_o$. Hence $w(A_n) \downarrow 0$.

If $\hat{\mu}$ is defined from a probability measure λ on a σ-field of sets $\mathcal{Q}$ as above, then the probability measure w defined on $\mathcal{Q}$ in theorem 3.2 is equal to λ. Note merely that $\lambda(A) = \sup\{\alpha : (\hat{\mu}([A]))(\alpha) = [X]\}$. Also, on the other hand, if $\hat{\mu}$ is a probability on $\mathcal{Q}/\lambda$ and w is a probability measure on $\mathcal{Q}$ as defined in theorem 3.2, then if $\hat{\nu}$ is a probability on $\mathcal{Q}/\lambda$ defined from w, then $\hat{\mu}$ and $\hat{\nu}$ are identical.

This section shows that there is no disadvantage in considering probabilities in the sense of section 2.2 and that there are numbers associated with these probabilities for one who must compute. However, there is the advantage of a simpler theory, a large part of which is in

Kappos. Another advantage brought out by this work is their relation to conditional probability as defined in section 2.4. This leads up to the question of the existence of numerical values properly associated with the conditional probabilities on a Boolean algebra which will be considered in the next section.

3.2. The Relation in Conditioning

Let $\mathcal{B}$ be a σ-subfield of $\mathcal{A}$ and let $\mathcal{B}/\lambda$ be the set of residue classes of $\mathcal{A}/\lambda$ which contain an element of $\mathcal{B}$ (v. section 1.3). The reason for this is to have $\mathcal{B}/\lambda \subset \mathcal{A}/\lambda$. As before, $\mathcal{B}/\lambda$ is a complete Boolean algebra.

A real-valued mapping $\lambda(\cdot,\cdot|\ \mathcal{A},\mathcal{B})$ defined on $\mathcal{A} \times X$ is a regular conditional probability on $\mathcal{A} \times X$ (v. Jirina [11]) if

(1) for any A in $\mathcal{A}$, $\lambda(A,\cdot|\mathcal{A},\mathcal{B})$ is $\mathcal{B}$-measurable, and

$$\lambda(A \cap B) = \int_B \lambda(A,x|\ \mathcal{A},\mathcal{B})d\lambda(x)$$

holds for all B in $\mathcal{B}$, and

(2) for every x in X, $\lambda(\cdot,x|\mathcal{A},\mathcal{B})$ is a probability measure on $\mathcal{A}$.

Condition (1) is the conditional probability condition and (2) is the regularity condition. Throughout the remainder of this work the term ''conditional probability on $\mathcal{A} \times X$'' will be used to distinguish this mapping from the conditional probability on a Boolean algebra.

A conditional probability on $\mathcal{A} \times X$, satisfies:

1. for every A in $\mathcal{A}$, $0 \leq \lambda(A,\cdot|\mathcal{A},\mathcal{B}) \leq 1$ with λ-probability one,
2. $\lambda(\phi,\cdot|\mathcal{A},\mathcal{B}) = 0$ with λ-probability one,

3. $\lambda(X,\cdot|\mathcal{A},\mathcal{B}) = 1$ with λ-probability one,

4. if $A \cap B = \phi$, $\lambda(A \cup B,\cdot|\mathcal{A},\mathcal{B}) = \lambda(A,\cdot|\mathcal{A},\mathcal{B})+\lambda(B,\cdot|\mathcal{A},\mathcal{B})$ with λ-probability one, and

5. if $A_n \downarrow \phi$, $\lambda(A_n,\cdot|\mathcal{A},\mathcal{B}) \downarrow 0$ with λ-probability one.

Define a mapping $\hat{\mu}_{\mathcal{B}/\lambda}$ on $\mathcal{A}/\lambda$ into $\Omega[\mathcal{B}/\lambda]$ by

$$(\hat{\mu}_{\mathcal{B}/\lambda}([A]))(\alpha) = [\{x:\lambda(A,x|\mathcal{A},\mathcal{B}) > \alpha\}].$$

The following theorem parallels theorem 3.1.

<u>Theorem 3.3</u>. If $\lambda(\cdot,\cdot|\mathcal{A},\mathcal{B})$ is a conditional probability on $\mathcal{A} \times X$, then $\hat{\mu}_{\mathcal{B}/\lambda}$ is a conditional probability.

<u>Proof</u>. Let $\lambda(\cdot,\cdot) = \lambda(\cdot,\cdot|\mathcal{A},\mathcal{B})$. In view of 1,2, and 3 for $\lambda(\cdot,\cdot)$ only the positivity, additivity, and continuity conditions will be checked.

If $\hat{\mu}_{\mathcal{B}/\lambda}([A]) = \hat{0}$, $\lambda(A,\cdot) = 0$ with λ-probability one, that is

$$\lambda(A \cap B) = \int_B \lambda(A,x)d\lambda(x) = 0, \quad \text{for every } B \text{ in } \mathcal{B}.$$

Thus $\lambda(A) = 0$ or, equivalently $[A] = [\phi]$.

If $[A] \wedge [B] = [\phi]$, choose A in $[A]$, B in $[B]$ so that $A \cap B = \phi$. Then

$$\begin{aligned}(\hat{\mu}_{\mathcal{B}/\lambda}([A] \vee [B]))(\alpha) &= [\{x:\lambda(A \cup B, x) > \alpha\}] \\ &= [\{x: \lambda(A,x) + \lambda(B,x) > \alpha\}] \\ &= [\bigcup_B(\{x:\lambda(A,x) > \beta\} \cap \{x:\lambda(B,x) > \alpha-\beta\})]\end{aligned}$$

$$= \bigvee_{\beta} \wedge([\{x:\lambda(A,x) > \beta\}], [\{x:\lambda(B,x) > \alpha-\beta\}])$$

$$= \bigvee_{\beta} \wedge((\hat{\mu}_{\mathcal{B}/\lambda}([A]))(\beta), (\hat{\mu}_{\mathcal{B}/\lambda}([B]))(\alpha-\beta))$$

$$= (\hat{\mu}_{\mathcal{B}/\lambda}([A]) + \hat{\mu}_{\mathcal{B}/\lambda}([B]))(\alpha).$$

If $[A_n] \downarrow [\phi]$ choose B_n in $[A_n]$ so that $B_n \downarrow \phi$ (for example, take $E_n = \bigcap_{k=1}^{n} A_k$ and use $B_n = E_n - E_\infty$). Given $\alpha > 0$, let

$$E_n = \{x:\lambda(B_n,x) > \alpha\}.$$

Then

$$\lambda(B_n \cap E_n) = \int_{E_n} \lambda(B_n,x)d\lambda_{|\mathcal{B}}(x) > \alpha\, \lambda(E_n),$$

that is

$$\lambda(B_n) > \alpha\, \lambda(E_n) \quad \text{so} \quad \lambda(E_n) < 1/\alpha\, \lambda(B_n) \downarrow 0.$$

Then $(\hat{\mu}_{\mathcal{B}/\lambda}([B_n]))(\alpha) = [E_n]$ where $[E_n] \downarrow [\phi]$, so $\bigwedge_n (\hat{\mu}_{\mathcal{B}/\lambda}([B_n]))(\alpha) = [\phi]$. Thus $\bigvee_{m=1}^{\infty} \bigwedge_n (\hat{\mu}_{\mathcal{B}/\lambda}([B_n]))(1/m) = [\phi]$, that is $(\bigwedge_n \hat{\mu}_{\mathcal{B}/\lambda}([B_n]))(0) = [\phi]$. In view of the fact that $\bigwedge_n \hat{\mu}_{\mathcal{B}/\lambda}([B_n]) \geq \hat{0}$, the proof is complete.

Notice that regularity is not involved in theorem 3.3. The next problem to be investigated is that of the existence of a regular conditional probability on $\mathcal{Q} \times X$ defined from a conditional probability in the sense of section 2.4. Of course, only a partial solution can be obtained (v. Dieudonne [4] or exercise 48(4) of Halmos [6]). However, in the cases important in application there is a solution. For the sake of completeness and intuitive appeal, the two most important of these cases will be included here.

In what follows a single set is chosen from each residue class of sets. Henceforth π denotes a mapping that does this, for example $\mathcal{B}/\lambda$ is taken into $\mathcal{B}$ by π. If f is a real-valued $\mathcal{B}$-measurable function on X, then $\hat{f}$ defined by

$$\hat{f}(\alpha) = [\{x:f(x) > \alpha\}]$$

is a random variable in the sense of 2.1 with values in $\mathcal{B}/\lambda$. If

$$g(x) = \sup\{\alpha:x \in \pi(\hat{f}(\alpha))\}$$

then g and f differ on a set of λ-probability zero. This illustrates the association between a $\mathcal{B}$-measurable function $\lambda(A,\cdot|\mathcal{Q},\mathcal{B})$ and a random variable $\hat{\mu}_{\mathcal{B}/\lambda}([A])$ which takes values in $\mathcal{B}/\lambda$. Therefore, if $\hat{\mu}_{\mathcal{B}/\lambda}$ is a conditional probability on $\mathcal{Q}/\lambda$ and $\lambda(\cdot,x)$ on $\mathcal{Q}$ is defined by

$$\lambda(A,x) = \sup\{\alpha: x \in \pi((\hat{\mu}_{\mathcal{B}/\lambda}([A]))(\alpha)\},$$

the work will be limited to showing $\lambda(\cdot,x)$ is a probability measure on $\mathcal{Q}$ for each x in X. When the set functions $\lambda(\cdot,x)$, x in X are probability measures, this class will be called a class of probability measures on $\mathcal{A}$ associated with $\hat{\mu}_{\mathcal{B}/\lambda}$. This definition of $\lambda(\cdot,\cdot)$ is used in the following theorem.

<u>Theorem 3.4</u>. Let $\mathcal{B}/\lambda$ be such that there is a mapping $\pi: \mathcal{B}/\lambda \to \mathcal{B}$ so that the class of images under π is a σ-field. Then the class $\{\lambda(\cdot,x): x \in X\}$ is a class of probability measures on $\mathcal{A}$ associated with $\hat{\mu}_{\mathcal{B}/\lambda}$.

<u>Proof</u>. Denote for convenience the subset $\pi((\hat{\mu}_{\mathcal{B}/\lambda}([A]))(\alpha))$ by $S(A,\alpha)$, then

$$\lambda(A,x) = \sup\{\alpha: x \in S(A,\alpha)\}.$$

It is evident that $0 \leq \lambda(A,x) \leq 1$ for all A in $\mathcal{A}$ and x in X. Also $\lambda(A,x) = 1$ for all x in X.

Let $A \cap B = \phi$, then by C.P.4

$$S(A \cup B,\alpha) = \bigcup_{\beta} (S(A,\beta) \cap S(B,\alpha-\beta)).$$

To show that $\lambda(A \cup B,x) = \lambda(A,x) + \lambda(B,x)$ we must show that

$$\sup\{\alpha: x \in \bigcup_{\beta}(S(A,\beta) \cap S(B,\alpha-\beta))\} = \sup\{\alpha: x \in S(A,\alpha)\} + \sup\{\alpha: x \in S(B,\alpha)\}.$$

If α_o is in the set of numbers on the left, there is a β_o so $x \in S(A,\beta_o)$ and also $S(B,\alpha_o-\beta_o)$, that is β_o is in the first set on the right and $\alpha_o-\beta_o$ is in the second. Thus the sum of the suprema on the right is larger than α_o and therefore larger than the supremum on the left. To show the other inequality suppose α_1 and α_2 are in the first and second sets on the right, respectively and that one of them say α_1 is not the supremum. Then there is, in the dense set of real numbers, a $\beta_o > \alpha_1$ such that $x \in S(A,\beta_o)$. Then also $x \in S(B,\alpha_2+\alpha_1-\beta_o)$ since $\alpha_2+\alpha_1-\beta_o<\alpha_2$. Consequently $x \in \bigcup_\beta (S(A,\beta) \cap S(B,\alpha_1+\alpha_2-\beta))$ so $\alpha_1+\alpha_2$ is in the set on the left.

Let $B_n \uparrow B$, then $S(B,\alpha) = \bigcup_{n=1}^{\infty} S(B_n,\alpha)$ (this follows similarly to the proof of C.P. 5 from C.E. properties). In order to show $\lambda(B) = \lim_n \lambda(B_n)$ we will show

$$\sup\{\alpha: x \in S(B,\alpha)\} = \lim_n \sup\{\alpha: x \in S(B_n,\alpha)\}.$$

If α is in the set on the left hand side, $x \in S(B,\alpha)$, that is there is n such that $x \in S(B_n,\alpha)$ so $\alpha \leq \lim_n \lambda(B_n,\alpha)$. Thus $\lambda(B,x) \leq \lim_n \lambda(B_n,x)$. The other inequality is obvious since $\lambda(B,x) \geq \lambda(B_n,x)$ for all n. Thus $\lambda(B,x) = \lim_n \lambda(B_n,x)$ and the proof is complete.

The problem of when such a π exists is open to investigation. In the discrete case there is such a π so the discrete case in probability is covered by theorem 3.4. The other important case that we will consider, summarized in theorem 3.5, will not be proved here because it is a special case of theorem 3.11. But, so that its intuitive value is not lost it will be given in the appendix.

<u>Theorem 3.5</u>. If X is the set of real numbers, $\mathcal{A}$ the Borel sets of X, $\mathcal{B}$ a σ-subfield of $\mathcal{A}$, and λ is a probability

on $\mathcal{Q}$, then there is a class of probability measures on $\mathcal{Q}$ associated with $\hat{\mu}_{\mathcal{B}/\lambda}$.

Theorems 3.4 and 3.5 are adequate for practical purposes, but theoretically the investigation must go farther. This leads to the consideration of perfect fields, compact measures, and perfect measures which will be taken up shortly. But first a useful tool will be introduced.

Although it is not necessary (see the proof of theorem 3.5 in the appendix), it is convenient to use a lifting, described by theorem 3 of Maharam [17]: If $(X, \mathcal{B}, \lambda)$ is complete, that is subsets of sets with λ-probability zero are events, a representative set $R(x) \subset X$ can be chosen for each class x of measurable sets modulo null sets in X, in such a way that $R(x)$ is measurable and in the measure class x, $R(0) = \phi$, $R(-x) = X-R(x)$, $R(x \wedge y) = R(x) \wedge R(y)$, and therefore, $R(x \vee y) = R(x) \vee R(y)$. For other results on liftings see Ionescu Tulcea ([8] and [9]).

There is no great loss in assuming completeness and we do it throughout the remainder of this work. Let π be a lifting on $\mathcal{B}/\lambda$ and define $\lambda(\cdot,x)$ on $\mathcal{Q}$ by

$$\lambda(A,x) = \sup\{\alpha : x \in \pi(\hat{\mu}_{\mathcal{B}/\lambda}([A]))(\alpha))\}$$

where $\hat{\mu}_{\mathcal{B}/\lambda}$ is a conditional probability in the sense of section 2.4.

A lifting differs from the mapping π of theorem 3.4 in that the set of images need not be a σ-field. Thus such sets as $\bigcup_{\beta}(S(A,\beta) \cap S(B,\alpha-\beta))$ need not be images under π but there is always an image under π which is λ-equivalent to such a set. Making appropriate changes in the proof of theorem 3.4 we have the following results.

Lemma 3.6. If $A \cap B = \phi$, then there is X_o in $\mathcal{B}$, $\lambda(X_o) = 0$ such that for x not in X_o $\lambda(A \cup B,x) = \lambda(A,x) + \lambda(B,x)$.

Corollary 3.7. If $\mathcal{Q}$ is a countable field, there is X_o in $\mathcal{B}$, $\lambda(X_o) = 0$ such that for x not in $X_o, \lambda(\cdot,x)$ is finitely additive on $\mathcal{Q}$.

Lemma 3.8. If $A_n \uparrow A$, then there is X_o in $\mathcal{B}, \lambda(X_o) = 0$ such that for x not in $X_o, \sup_n \lambda(A_n,x) = \lambda(A,x)$.

In theorem 2.1 we used the fact that the field $\mathcal{E}$ was perfect to make an extension, this brings up the question of the connection between perfect fields of sets and compact classes which is studied briefly below.

Theorem 3.9. If $\mathcal{B}$ is a perfect field of subsets of X, $\mathcal{B}$ is a compact class.

Proof. Let B_n be a sequence of $\mathcal{B}$ such that $\bigcap_{i=1}^{n} B_i \neq \phi$, n=1,2,... and suppose $\bigcap_{n=1}^{\infty} B_n = \phi$, then

$$X = \bigcup_{n=1}^{\infty} B_n^c = B_1^c \cup (B_1 \cap B_2^c) \cup (B_1 \cap B_2 \cap B_3^c) \cup \cdots$$

where the components of the union are disjoint. This means there is n_o such that $n > n_o$ implies $B_1 \cap \cdots \cap B_n \cap B_{n+1}^c = \phi$ (v. Sikorski [23]).

Then $B_{n+1} \supset B_1 \cap \dots \cap B_n$ and so $B_1 \cap \dots \cap B_n = (B_1 \cap \dots \cap B_n) \cap B_{n+1} \subset B_{n+2}$, inductively $B_{n+j} \supset B_1 \cap \dots \cap B_n$, $n > n_o$ and j. Hence

$\bigcap_{n=1}^{\infty} B_n = B_1 \cap \dots \cap B_n = \phi$ in contradiction with our original assumption.

Theorem 3.10 is a partial converse to theorem 3.9.

Theorem 3.10. If $\mathcal{B}$ is a countable field of subsets of a set X with the property that $\mathcal{B}$ is a compact class, $\mathcal{B}$ is a perfect field of subsets of X.

Proof. Let $\mathcal{B}$ be an open basis for a topology on X, then the sets of $\mathcal{B}$ are both open and closed. Let G be an open cover of X and assume the sets of G belong to $\mathcal{B}$. Since $\mathcal{B}$ is countable there is a sequence B_n of $\mathcal{B}$ for which

$$X = \bigcup_{n=1}^{\infty} B_n = B_1 \cup (B_1^c \cap B_2) \cup (B_1^c \cap B_2^c \cap B_3) \cup \dots$$

where each component of the union is in $\mathcal{B}$. Then since $\mathcal{B}$ is a compact class,

$$B_1^c \cap (B_1 \cup B_2^c) \cap (B_1 \cup B_2 \cup B_3^c) \cap \dots = \phi$$

implies that

$$B_1^c \cap (B_1 \cup B_2^c) \cap \dots \cap (B_1 \cup \dots \cup B_{n-1} \cup B_n^c) = \phi$$

for some n. It follows that $\bigcap_{j=1}^{n} B_j^c = \phi$, hence $X = \bigcup_{j=1}^{n} B_j$, that is X is compact.

Let Δ be a maximal ideal in $\mathcal{B}$, $\Delta = \{B_1, B_2, \ldots\}$. If $\bigcup_{n=1}^{\infty} B_n = X$, $X = B_{n_1} \cup \cdots \cup B_{n_j} \in \Delta$, in contradiction to Δ maximal. Hence there is a point x in $X - \bigcup_{n=1}^{\infty} B_n$. Let Δ_x be the maximal ideal determined by x. Since each $B_n \in \Delta_x$, $\Delta \subset \Delta_x$ and thus any maximal ideal in $\mathcal{B}$ is generated by a point.

Return now to the problem of classes of probability measures associated with a conditional probability on a Boolean algebra. The theorem that follows is not new (v. Sazonov [21]), but the proof is different and more natural in the present setting.

Let $(X, \mathcal{Q}, \lambda)$ be as in the preceding sections. Let $\mathcal{B}_1$ be a subfield of $\mathcal{Q}$, $\mathcal{B}_2$ a σ-subfield of $\mathcal{Q}$ for which every subset of a λ-null set in $\mathcal{B}_2$ is in $\mathcal{B}_2$. Then exactly as in section 2.4, there is a conditional probability $\hat{\mu}_{\mathcal{B}_2/\lambda}$ on $\mathcal{B}(\mathcal{B}_1)/\lambda$ into $\Omega[\mathcal{B}_2/\lambda]$, where $\mathcal{B}(\mathcal{B}_1)$ is the σ-field generated by $\mathcal{B}_1$.

> <u>Theorem 3.11</u>. If $\hat{\mu}_{\mathcal{B}_2/\lambda}$ is a conditional probability on $\mathcal{B}(\mathcal{B}_1)/\lambda$, $\mathcal{B}_1$ is a countable field, and $\lambda|_{\mathcal{B}_1}$ is compact, then there is a class of probability measures on $\mathcal{B}(\mathcal{B}_1)$ associated with $\hat{\mu}_{\mathcal{B}_2/\lambda}$.

<u>Proof</u>. If π is a lifting on $\mathcal{B}_2/\lambda$, define $\lambda(B,x) = \sup\{\alpha : x \in \pi((\hat{\mu}_{\mathcal{B}_2/\lambda}([B]))(\alpha))\}$ where B is in $\mathcal{B}_1$. By corollary 3.7, $\lambda(\cdot, x)$ is finitely additive for all x outside a λ-null set. Since $\lambda|_{\mathcal{B}_1}$ is compact, there is a compact class $\mathcal{C}$ such that

for every $B \in \mathcal{B}_1$ there is C_n in $\mathcal{C}$ and B_n in $\mathcal{B}_1$ such that

$$B_n \subset C_n \subset B \quad \text{and} \quad \lambda(B) = \sup_n \lambda(B_n).$$

By lemma 3.8, for all x outside a λ-null set,

$$\lambda(B,x) = \sup_n \lambda(B_n,x).$$

$\mathcal{B}_1$ countable implies that $\lambda(\cdot,x)$ is also compact for x not in a λ-null set. Let $N \subset \mathcal{B}_2$ be a λ-null set for which $\lambda(\cdot,x)$ is a compact, finitely additive measure on $\mathcal{B}_1$ for x not in N. For B in $\mathcal{B}_1$, define

$$\mu(A,x) = \begin{cases} \lambda(A,x), & x \notin N \\ \lambda(A), & x \in N . \end{cases}$$

But $\mu(\cdot,x)$ can be extended to a probability on $\mathcal{G}(\mathcal{B}_1)$, for each x (v. Marczewski [18]). The proof is complete.

Actually the family $\{\mu(\cdot,x):x \in X\}$ is equi-compact on $\mathcal{B}_1$. A family $\{\mu(\cdot,x):x \in X\}$ of finitely additive probability measures on a field $\mathcal{B}$ of subsets of X is equi-compact if there is a compact class $\mathcal{C}$ of subsets such that for every A in $\mathcal{B}$ and for every $\eta > 0$ there is B in $\mathcal{B}$ and C in $\mathcal{C}$ such that $B \subset C \subset A$ and $\mu(A-B ,x) < \eta$ for every x in X. With this it can be seen that if $\lambda(\cdot,\cdot|\mathcal{G}(\mathcal{B}_1),\mathcal{B}_2)$ is a regular conditional probability on $\mathcal{G}(\mathcal{B}_1) \times X$ and $\mathcal{B}_1$ is an arbitrary field and if $\lambda(\cdot,x|\mathcal{B}_1,\mathcal{B}_2)$ is an equi-compact family, then $\lambda|\mathcal{B}_1$ is compact. But this is outside the scope of this work.

The problem of finding probability measures associated with a conditional probability on a Boolean algebra of measurable sets modulo the null sets is equivalent to finding a regular conditional probability on the product of the measurable sets with the point space in the conventional theory. This problem and its converse has been studied for years, it is hopeful that the present work may be of assistance in the solution of this problem.

SUMMARY AND CONCLUSIONS

This work is divided into three chapters. In chapter 1 some of the properties of the Stone space of a Boolean algebra are adapted to the problem of conditioning and it is shown (theorem 1.4) that there is always a regular conditional probability relative to a full σ-subalgebra of Baire sets.

Chapter 2 contains a treatment of conditional probability on a Boolean algebra. For this a generalized integral is defined (section 2.2) and the standard theory of integration is shown to hold for it. A new defintion of probability is given (section 2.2), and conditional expectation and conditional probability are defined (section 2.4). The properties of conditional expectations and conditional probability are given in theorems 2.12 and 2.13, respectively. It is pointed out that there is no regularity condition for conditional probability on a Boolean algebra. Integrals in section 2.2 are defined with respect to certain mappings, an example of which is conditional probability and the conditional expectation of a random variable is shown to be the integral of the random variable with respect to conditional probability (theorem 2.14).

The relation of the theory developed in chapter 2 to the conventional theory is presented in the third chapter. One of the aspects of this is to show there are numerical values properly associated with probabilities and conditional probabilities as defined in this work, and it is done in chapter 3. The relation of compact classes to perfect fields of sets is also discussed briefly.

ACKNOWLEDGEMENT

The author would like to thank Professor Baxter and the Air Force Office of Scientific Research for their financial assistance and to Professor Cote for his advice and encouragement.

BIBLIOGRAPHY

1. Anderson, E.S. and B. Jessen (1948). On the introduction of measures in infinite product sets. Danske Vid. Selsk. Math.-Fys. Medd., v. 25, pp. 1-8.

2. Birkhoff, G. (1961). Lattice Theory. Amer. Math. Soc. Colloq. Publ., V. XXV, Providence.

3. Bourbaki, N. (1952-1963). Integration. Hermann, v. I-VIII, Paris.

4. Dieudonne, J. (1948). Sur le theoreme de Lebesgue-Nikodym, III. Ann. Univ. Grenoble, v. 23, pp. 25-53.

5. Doob, J. (1953). Stochastic Processes. Wiley, New York.

6. Halmos, P. (1950). Measure Theory. Van Nostrand, New York.

7. Halmos, P. (1949). On a theorem of Dieudonne. Proc. Nat. Acad. Sci. U.S.A., v. 35, pp. 38-42.

8. Ionescu Tulcea, C. (1965). On the lifting property and disintegration of measures. Bull. Amer. Math. Soc., v. 71, pp. 829-842.

9. Ionescu Tulcea, A. and C. Ionescu Tulcea (1961). On the lifting property I. J. Math. Anal. Appl., v. 3, pp. 537-546.

10. Jessen, B. (1948). On two notions of independent functions. Colloq. Math., v. 1, pp. 214-215.

11. Jirina, M. (1954). Conditional probabilities on σ-algebras with countable basis. Selected Translations in Mathematical Statistics and Probability, v. 2, pp. 79-85.

12. Kappos, D. (1960). Strukturtheorie der Wahrscheinlichkeits-felder und Raume. Ergebnisse der Mathematik. Springer, Berlin.

13. Kelley, J.L. (1955). General Topology. Van Nostrand, New York.

14. Kolmogorov, A. (1933). Grundbegriffe der Wahrscheinlichkeitrechnung Ergebnisse der Mathematik, Springer, Berlin. (1956). Foundations of the Theory of Probability. Chelsea, New York.

15. Kolmogorov, A. (1948). Algebres de Boole metriques completes. VI Zjazd Matematykow Pol., Dodatek do Rocznika Pol. Tow. Mat., v. 20, pp. 21-30.

16. Loeve, M. (1960). Probability Theory. Van Nostrand, New York.

17. Maharam, D. (1958). On a theorem of von Neumann. Proc. Amer. Math. Soc., v. 9, pp. 987-994.

18. Marczewski, E. (1953). On compact measures. Fund. Math., v. 40, pp. 111-124.

19. Olmsted, J.M.H. (1942). Lebesgue theory on a Boolean algebra. Trans. Amer. Math. Soc., v. 51, pp. 164-193.

20. Ryll-Nardzewski, C. (1953). On quasi-compact measures. Fund. Math., v. 40, 125-130.

21. Sazonov, V. (1962). On perfect measures. Amer. Math. Soc. Transl., v. 48, pp. 229-254.

22. Segal, I. (1965). Algebraic integration theory. Bull. Amer. Math. Soc., v. 71, pp. 419-489.

23. Sikorski, R. (1964). Boolean Algebras. Ergebnisse der Mathematik. Springer, Berlin.

24. Varadarajan, V.S. (1962). Probability in Physics and a theorem on simultaneous observability. Comm. Pure Appl. Math., v. 15, pp. 189-217.

APPENDIX

A Proof of Theorem 3.5

This proof is a combination of ideas used in theorem 9.4 of Doob [5] and prop. 3.3 of Varadarajan [24], adapted to the setting of the present work. The proof will be preceded by the following lemma.

Lemma. If $r_1, r_2, \ldots$ is a distinct enumeration of the rationals in the interval $(0,1)$, λ is any real number, and if there is a class of subsets $B(\lambda, r_i)$ of a set X such that $B(\lambda, r_i) \supset B(\lambda, r_j)$ if $r_i < r_j$ and

i) $B(\lambda, r_i) \subset B(\mu, r_i)$ if $\lambda < \mu$,

ii) $B(\lambda, r_i) = \bigcup_n B(\lambda_n, r_i)$ if $\lambda_n \uparrow \lambda$,

iii) $X = \bigcup_n B(\lambda_n, r_i)$ if $\lambda_n \uparrow \infty$, and $\phi = \bigcap_n B(\lambda_n, r_i)$ if $\lambda_n \downarrow -\infty$

for any r_i, then for every x in X there is a probability measure on the Borel sets of the real line given by $F(\lambda) = \sup\{r_i : x \in B(\lambda, r_i)\}$.

Proof. Let us consider a particular point x.

1) If $\lambda < \mu$, $B(\lambda, r_i) \subset B(\mu, r_i)$ for all r_i, that is $F(\lambda) = \sup\{r_i : x \in B(\lambda, r_i)\} \leq \sup\{\alpha : x \in B(\mu, r_i)\} = F(\mu)$.

2) If $\lambda_n \uparrow \lambda$, $B(\lambda, r_i) = \bigcup_n B(\lambda_n, r_i)$ for all r_i, that is for each r_i for which $x \in B(\lambda, r_i)$, $x \in B(\lambda_n, r_i)$ for some n, or equivalently, $r_i \leq F(\lambda_n)$ for some n. Thus $F(\lambda) \leq \sup_n F(\lambda_n)$. Moreover, equality

prevails since the reverse inequality is obvious from 1).

3) If $\lambda_n \uparrow \infty$, $X = \bigcup_n B(\lambda_n, r_i)$ for all r_i, that is for any r_i there is n_o such that $n > n_o$ implies $r_i \leq F(\lambda_n) \leq 1$ so that $\sup_n F(\lambda_n)=1$. Similarly if $\lambda_n \downarrow -\infty$, $\phi = \bigcap_n B(\lambda_n, r_i)$ for all r_i, that is for any r_i there is n for which $x \notin B(\lambda_n, r_i)$. But $B(\lambda_n, r_i) \supset B(\lambda_n, r_j)$ for all $r_i < r_j$ so $0 \leq F(\lambda_n) \leq r_i$. Then $\inf_n F(\lambda_n) = 0$.

From 1),2), and 3) above it is clear that $F(\lambda)$ is a distribution function for each x in X. Each distribution function generates a probability measure on the Borel sets of the real line.

Notice that there is no restriction on X in the proof. This lemma corresponds to the wide sense notion and the proof of theorem 9.4 of [5].

Theorem 3.5. If X is the set of real numbers, $\mathcal{A}$ the Borel sets of X, $\mathcal{B}$ a σ-subfield of $\mathcal{A}$, and λ is a probability on $\mathcal{A}$, then there is a class of probability measures on $\mathcal{A}$ associated with $\hat{\mu}_{\mathcal{B}/\lambda} : \mathcal{A}|\lambda \to \Omega[B/\lambda]$.

Proof. (cf. Varadarajan [24]). For $r \leq 0$ or $1 \leq r$, let $B(\lambda, r) = X$ or ϕ, respectively where λ is any rational. Let $r_1, r_2, \ldots$ be as in the lemma. By induction we will show that there are sets $B(\lambda, r_1)$, $B(\lambda, r_2), \ldots$ in $\mathcal{B}$ such that i) $(\hat{\mu}_{\mathcal{B}/\lambda}([(-\infty, \lambda]]))(r_j) = [B(\lambda, r_j)]$, and ii) $B(\lambda, r_i) \supset B(\lambda, r_j)$ if $r_i < r_j$. For convenience let $\hat{\mu}(\lambda) = \hat{\mu}_{\mathcal{B}/\lambda}([(-\infty, \lambda]])$. Let $B(\lambda, r_1)$ be any set for which $(\hat{\mu}(\lambda)(r_1) = [B(\lambda, r_1)]$. If $B(\lambda, r_1), \ldots, B(\lambda, r_n)$ have been constructed to fit i) and ii), $B(\lambda, r_{n+1})$ is constructed as follows. Let $r_{i_1} < \dots < r_{i_n}$ be the arrangement of $r_1, \dots, r_n$ in increasing order. Choose B such that $(\hat{\mu}(\lambda))(r_{n+1}) = [B]$. There are three possibilities 1) $r_{n+1} < r_{i_1}$, in which case let $B(\lambda, r_{n+1}) = B \cup B(\lambda, r_{i_1})$, 2) $r_{i_\mu} < r_{n+1} < r_{i_{\mu+1}}$, some $\mu = 1, \dots, n-1$, in which case let $B(\lambda, r_{n+1}) = B(\lambda, r_{i_\mu}) \cup (B(\lambda, r_{i_{\mu+1}}) \cap B)$, and 3) $r_{i_n} < r_{n+1}$, in which case let $B(\lambda, r_{n+1}) = B \cap B(\lambda, r_{i_n})$. Then $(\hat{\mu}(\lambda))(r_{n+1}) = [B(\lambda, r_{n+1})]$ in each case. By induction the sets $B(\lambda, r_1)$, $B(\lambda, r_2), \dots$ are found satisfying i) and ii).

Let $C = (\bigcap_i \bigcup_\lambda B(\lambda, r_i))^c$ and $D = \bigcup_i \bigcap_\lambda B(\lambda, r_i)$. In what follows the properties C.P.1-C.P.5 of $\hat{\mu}_{\mathcal{B}/\lambda}$ are often used and will not be explicitly stated. Note $[X] = (\hat{\mu}(\infty))(r_i) = \bigvee_\lambda (\hat{\mu}(\lambda))(r_i)$. But $(\hat{\mu}(\lambda))(r_i) = [B(\lambda, r_i)]$ so $[X] = \bigvee_\lambda [B(\lambda, r_i)]$, and

$[\phi] = (\hat{\mu}(-\infty))(r_i) = \bigvee_{m=1}^{\infty} \bigwedge_\lambda (\hat{\mu}(\lambda))(r_i + 1/m)$ so if r_j is given there is $0 < r_i + 1/m < r_j$, consequently $[\phi] = \bigwedge_\lambda (\hat{\mu}(\lambda))(r_i + 1/m) \geq \bigwedge_\lambda (\hat{\mu}(\lambda))(r_j)$ $= \bigwedge_\lambda [B(\lambda, r_j)] = [\bigcap_\lambda B(\lambda, r_j)]$. Therefore $[C] = [D] = [\phi]$.

Let

$$A(\mu, r_i) = \begin{cases} \bigcup_{\lambda < \mu} (B(\lambda, r_i) \cup C), & \mu > 0 \\ \bigcup_{\lambda < \mu} (B(\lambda, r_i) \cap D^c), & \mu \leq 0. \end{cases}$$

Then, if $\mu > 0$, $[A(\mu,r_i)] = \bigvee_{\lambda < \mu} [B(\lambda,r_i) \cup C] = \bigvee_{\lambda < \mu} [B(\lambda,r_i)]$

$= \bigvee_{\lambda < \mu} (\hat{\mu}(\lambda))(r_i) = (\hat{\mu}(\mu))(r_j) = [B(\mu,r_i)]$. If $\mu \leq 0$, $[A(\mu,r_i)] =$

$\bigvee_{\lambda < \mu} [B(\lambda,r_i) \cap D^c] = \bigvee_{\lambda < \mu} [B(\lambda,r_i)] = [B(\mu,r_i)]$.

For $r_i < r_j$, and if $\mu > 0$, $A(\mu,r_i) \cup A(\mu,r_j) = (\bigcup_{\lambda < \mu} B(\lambda,r_i)) \cup$

$(\bigcup_{\nu < \mu} B(\nu,r_j)) \cup C = \bigcup_{\lambda < \mu} (B(\lambda,r_i) \cup B(\lambda,r_j)) \cup C = \bigcup_{\lambda < \mu} B(\lambda,r_j) \cup C = A(\mu,r_j)$,

and if $\mu \leq 0$, $A(\mu,r_i) \cup A(\mu,r_j) = \bigcup_{\lambda < \mu} (B(\lambda,r_i) \cup B(\lambda,r_j)) \cap D^c$

$= \bigcup_{\lambda < \mu} B(\lambda,r_j) \cap D^c = A(\mu,r_j)$. Thus $A(\mu,r_i) \supset A(\mu,r_j)$ for $r_i < r_j$.

For r_i fixed, $\lambda < \mu$, then if $0 < \lambda$, $A(\lambda,r_i) = \bigcup_{\nu < \lambda} (B(\nu,r_i) \cup C) \subset$

$\bigcup_{\nu < \lambda} (B(\nu,r_i) \cup C) = A(\mu,r_i)$, and if $\lambda \leq 0$, $A(\lambda,r_i) =$

$\bigcup_{\nu < \lambda} (B(\nu,r_i) \cap D^c) \subset \bigcup_{\nu < \mu} (B(\nu,r_i) \cap D^c) \subset \bigcup_{\nu < \mu} (B(\nu,r_i) \cup C)$, but

$A(\mu,r_i) = \bigcup_{\nu < \mu} (B(\nu,r_i) \cap D^c)$ or $\bigcup_{\nu < \mu} (B(\nu,r_i) \cup C)$ depending on

whether $\mu \leq 0$ or $\mu > 0$. Thus $A(\lambda,r_i) \subset A(\mu,r_i)$ when $\lambda < \mu$.

Let $\mu_n \uparrow \mu$, then if $\mu \leq 0$, $\bigcup_n A(\mu_n,r_i) = \bigcup_n (\bigcup_{\lambda < \mu_n} (B(\lambda,r_i) \cap D^c)) =$

$\bigcup_{\lambda < \mu} (B(\lambda,r_i) \cap D^c) = A(\mu,r_i)$, and if $\mu > 0$, $\bigcup_n A(\mu_n,r_i) =$

$(\bigcup_{n:\mu_n > 0} A(\mu_n,r_i)) \cup (\bigcup_{n:\mu_n \leq 0} A(\mu_n,r_i)) = \bigcup_{n.\mu_n > 0} A(\mu_n,r_i) =$

$\bigcup_{n:\mu_n > 0} (\bigcup_{\lambda < \mu_n} (B(\lambda,r_i) \cup C)) = \bigcup_{\lambda < \mu} (B(\lambda,r_i) \cup C) = A(\mu,r_i)$. There-

fore $\bigcup_n A(\mu_n,r_i) = A(\mu,r_i)$ where $\mu_n \uparrow \mu$.

Note also that $\lim_{\mu\to\infty} A(\mu,r_i) = \bigcup_\lambda (B(\lambda,r_i) \cup C) = \bigcup_\lambda B(\lambda,r_i) \cup (\bigcap_i \bigcup_\lambda B(\lambda,r_i))^c$

$\supset \bigcup_\lambda B(\lambda,r_i) \cup (\bigcup_\lambda B(\lambda,r_i)))^c = X$, and $\lim_{\mu\to-\infty} A(\mu,r_i) = \bigcap_\lambda (B(\lambda,r_i) \cap D^c) =$

$\bigcap_\lambda B(\lambda,r_i) \cap (\bigcup_i \bigcap_\lambda B(\lambda,r_i))^c \subset \bigcap_\lambda B(\lambda,r_i) \cap (\bigcap_\lambda B(\lambda,r_i))^c = \phi.$

If μ is irrational let $A(\mu,r_i) = \bigcup_{\lambda < \mu} A(\lambda,r_i)$, where λ is rational. Then the sets $A(\mu,r_i)$ have the properties listed in the lemma. The proof is terminated with the application of the lemma.