ISBN 82-553-0437-1

No 1
1981

SUFFICIENT ALGEBRA'S OF EVENTS

Notes to Erik Torgersen's lectures
on decision theory, 1980

Recorded by Terje Aven

STATISTICAL MEMOIRS
INSTITUTE OF MATHEMATICS
UNIVERSITY OF OSLO

Sufficient σ-algebra's of events.

There are several interesting ways of introducing the concept of sufficiency. The most usual one says, roughly, that a statistic S is sufficient if conditional probabilities given S are themselves statistics, i.e. does not depend on the unknown parameter.

That sufficient statistics really are "sufficient" for statistical decision problems may then be argued by saying that the original experiment may be recovered from S by using a known random mechanism, i.e. the conditional distribution given S.

Another non-Bayesian way of considering sufficiency is in terms of operational characteristics. Thus we may say that S is sufficient provided there to each decision rule corresponds another depending on S only having the same operational characteristics.

Closely related is the "risk definition" of sufficiency which says that S is sufficient provided every obtainable riskfunction is dominated by a riskfunction obtainable from S.

According to Le Cam's theory on approximate sufficiency this amounts to say that the experiment defined by S is 0-deficient w.r.t. the given experiment.

Finally we would also like to mention that sufficiency may be introduced within a Bayesian context by saying that S is sufficient provided the unknown parameter and the set of observations are independent given S.

Other and equivalent ways of defining conditional independence (i.e. the Markov property) leads to equivalent definition of Bayesian sufficiency. Thus we may say that S is sufficient if the posterior distribution depends on S only.

An exposition of some results in this direction may be found in Torgersen (1976).

We shall here restrict ourselves to give an exposition of some of the basic machinery related to the first definition.

The reader who wants a deeper understanding of this theory cannot avoid to study the fundamental treatise by Wald (1950) on decision theory and the original papers by Bahadur (1954, 1955), Burkholder (1961), Dynkin (1951), Halmos and Savage (1949). He might also consult Le Cam (1964), Heyer (1973) and Torgersen (1975).

We have completely avoided the problem of the relationship between the concepts of sufficiency of statistics and sufficiency of σ-algebra's of events.

The theory presented here is adequate provided a sufficient statistic is defined as a statistic whose induced σ-algebra is sufficient. Thus we do not consider all measurable inverse images of sets, only inverse images of measurable sets. The reader may consult the paper by Landers (1974) and the references there to get an impression of the difficulties which may occure.

Let us finally comment that the "risk definition" of sufficiency as described in Le Cam (1964) is equivalent to pairwise conditional expectation sufficiency.

A most thorough treatment of pairwise sufficiency may be found in Siebert (1979).

Definition 1

Let (X,A,P) be a probability space and let $\mathcal{B}$ be a sub-σ-algebra of A. If X is an extended real valued random variable (r.v.) such that EX exists, then there is a unique (a.e.(P)) $\mathcal{B}$-measurable function $E(X|\mathcal{B})$ such that

$$\int_B E(X|\mathcal{B})dP = \int_B XdP \qquad \text{for all } B\in\mathcal{B}$$

A few useful properties of conditional expectations:

(i) $X \geq 0$ a.e. $\Rightarrow$ $E(X|\mathcal{B}) \geq 0$ a.e.

(ii) $E[EX|\mathcal{B}]^{\pm} \leq EX^{\pm}$

(iii) $E(EX|\mathcal{B}) = EX$

(iv) If $E|X_i|<\infty$ $i=1,2$ and Y_i $i=1,2$ are $\mathcal{B}$ measurable and bounded then

$$E(Y_1X_1+Y_2X_2|\mathcal{B}) = Y_1E(X_1|\mathcal{B}) + Y_2E(X_2|\mathcal{B}) \quad \text{a.e.}$$

Definition 2

Let $\xi = (X,A;P_\theta : \theta\in\Theta)$ be an experiment, i.e. (X,A) is a measure space and $\{P_\theta,\theta\in\Theta\}$ is a family of probability measures on (X,A). Then a sub-σ-algebra $\mathcal{B}$ of A is said to be sufficient for ξ if corresponding to each $A\in A$, there exists a $\mathcal{B}$-measurable function Y_A such that

$$P_\theta(A|\mathcal{B}) = Y_A \qquad \text{a.e.} \quad (P_\theta)\ \forall\theta\in\Theta$$

This condition is equivalent to the following:
To each bounded or non-negative $\mathcal{A}$-measurable real function Z on χ there corresponds a $\mathcal{B}$-measurable function Y_Z such that

$$E_\theta(Z|\mathcal{B}) = Y_Z \qquad \text{a.e.} \quad (P_\theta)\ \forall\theta\in\Theta$$

Thus the conditional probability of an event A (respectively conditional expectation of a random variable Z) may be specified (almost) independent of the parameter $\theta\in\Theta$.

Definition 3

Two measures μ and ν on a measurable space $(\chi,\mathcal{A})$ are said to be equivalent if $\nu<<\mu$ and $\mu<<\nu$ (i.e. ν is absolutely continuous w.r.t. μ and vice versa). We then write $\mu\sim\nu$.
The relation $\sim$ is an equivalence relation on the set of measures on $(\chi,\mathcal{A})$, and the equivalence classes consist of measures having the same null-sets.

Lemma 4

Let μ be a σ-finite measure on the measurable space $(\chi,\mathcal{A})$, $\mu\not\equiv 0$. Then there is a probability measure P on $(\chi,\mathcal{A})$ such that $\mu\sim P$.

Proof

If μ is finite, then define $P(A) = \dfrac{\mu(A)}{\mu(\chi)}$ for each $A\in\mathcal{A}$.
P is easily seen to have the required properties. Assume μ is σ-finite but not finite. Then we may write $\chi = \bigcup_{i=1}^{\infty} \chi_i$ where $\chi_i\in\mathcal{A}$ and $0<\mu(\chi_i)<\infty$. Replacing $(\chi_1,\chi_2,\chi_3,\ldots)$ with $(\chi_1,\chi_2-\chi_1,\chi_3-(\chi_1\cup\chi_2),\ldots)$ we see that we may assume that the sets $\chi_1,\chi_2,\ldots$ are disjoint.

For each $A \in \mathcal{A}$ define $P(A) = \sum_{n=1}^{\infty} a_n \frac{\mu(A \cap \chi_n)}{\mu(\chi_n)}$ where

$a_n > 0 \quad \forall n$ and $\sum_{n=1}^{\infty} a_n = 1$

Clearly P is a probability measure on $(\chi, \mathcal{A})$ since μ is a measure on $(\chi, \mathcal{A})$ and $P(\chi) = \sum_{n=1}^{\infty} a_n = 1$

Now $P(A) = 0 \Leftrightarrow \mu(A \cap \chi_n) = 0 \quad \forall n \Leftrightarrow \mu(A) = \sum_{n=1}^{\infty} \mu(A \cap \chi_n) = 0$

Thus $P \sim \mu$.

Lemma 5

Let $(\chi, \mathcal{A}, \mu)$ be a σ-finite measure space and X and X' $\mathcal{A}$-measurable functions on χ .

Assume $\int_A X d\mu = \int_A X' d\mu$ for all $A \in \mathcal{A}$.

Then $X = X'$ a.e. (μ)

Proof

Since μ is σ-finite, we may write $\chi = \bigcup_{i=1}^{\infty} \chi_i$ where

$\chi_i \in \mathcal{A}$, $\chi_i \cap \chi_j = \emptyset \quad i \neq j$ and $\mu(\chi_i) < \infty$.

Let $A = \{x \in \chi \mid X(x) > X'(x)\}$. Then $A \cap \chi_i \in \mathcal{A}$ for all i.

Now we must have $\mu(A \cap \chi_i) = 0 \quad \forall i$. For if $\mu(A \cap \chi_i) > 0$, then

$$\int_{A \cap \chi_i} X d\mu > \int_{A \cap \chi_i} X' d\mu$$

Thus $\mu(A) = \sum_{i=1}^{\infty} \mu(A \cap \chi_i) = 0$

By symmetry $\mu\{X' > X\} = 0$ and hence $\mu\{X \neq X'\} = 0$

Theorem 6

Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space and $\mathcal{B}$ a sub-σ-algebra of $\mathcal{A}$. Furthermore let Z be a $\mathcal{A}$-measurable, non-negative real function on X. Then there exists a unique (a.e.(μ)) $\mathcal{B}$-measurable, non-negative function $\mu(Z|\mathcal{B})$ such that

$$(1.1) \qquad \int_B \mu(Z|\mathcal{B})d\mu = \int_B Z d\mu \qquad \text{for all } B \in \mathcal{B} .$$

Proof

By Lemma 4 there exists a probability measure P defined on $(X, \mathcal{A})$ such that $P \sim \mu$, and by Radon-Nikodym there is a non-negative $\mathcal{A}$-measurable function $h = \frac{d\mu}{dP}$ such that $\mu(A) = \int_A h dP$ for each A in $\mathcal{A}$.

Thus (1.1) is equivalent to

$$\int_B \mu(Z|\mathcal{B})h dP = \int_B Z h dP \qquad \text{for all } B \in \mathcal{B}$$

(see the proof of Lemma 8) which, in its turn, is equivalent to

$$\int_B E[\mu(Z|\mathcal{B})h|\mathcal{B}]dP = \int_B E(Zh|\mathcal{B})dP \qquad \text{for all } B \in \mathcal{B}$$

by definition 1. Since $\mu(Z|\mathcal{B})$ is required to be $\mathcal{B}$-measurable the left side equals $\int_B \mu(Z|\mathcal{B})E(h|\mathcal{B})dP$.

Note that $E(Zh|\mathcal{B})$ and $E(h|\mathcal{B})$ are $\mathcal{B}$-measurable and non-negative. Thus by Lemma 5, (1.1) is equivalent to

$$\mu(Z|\mathcal{B})E(h|\mathcal{B}) = E(Zh|\mathcal{B}) \quad \text{a.e.} \quad (P)$$

Since

$$\mu\{E(h|\mathcal{B}) = 0\} = \int_{\{E(h|\mathcal{B})=0\}} h dP = \int_{\{E(h|\mathcal{B})=0\}} E(h|\mathcal{B})dP = 0$$

we can define

$$\mu(Z|\mathcal{B}) = \frac{E(Zh|\mathcal{B})}{E(h|\mathcal{B})} \quad \text{a.e.} \quad (\mu)$$

The above arguments also show that $\mu(Z|\mathcal{B})$ is unique (a.e. (μ)).

Corollary 7

Let (X, A, μ) be a σ-finite measure space and B a sub-σ-algebra of A.

Then there exists a unique (a.e. (μ)) B-measurable, non-negative function $\mu(A|B)$ such that

$$\int_B \mu(A|B)d\mu = \int_B I_A d\mu = \mu(A \cap B) \quad \text{for all} \quad B \in B .$$

Proof

Let $Z = I_A$ in Theorem 6 and write $\mu(I_A|B) = \mu(A|B)$

Lemma 8

Let λ, γ, ν be σ-finite measures on (X, A) such that

$$\nu << \gamma << \lambda \quad .$$

Then $\frac{d\nu}{d\lambda} = \frac{d\nu}{d\gamma}\frac{d\gamma}{d\lambda}$ a.e. (λ)

Proof

Since $\gamma << \lambda$, we must have

$$(1.2) \qquad \int f d\gamma = \int f \frac{d\gamma}{d\lambda} d\lambda$$

when f is indicator function. Hence it follows by standard extension to simple functions and monotone limits of simple functions that (1.2) holds for every non-negative measurable function f.

Letting $f = I_A \frac{d\nu}{d\gamma}$, we get

$$\int_A \frac{d\nu}{d\gamma} d\gamma = \int_A \frac{d\nu}{d\gamma}\frac{d\gamma}{d\lambda} d\lambda \quad \text{for every} \quad A \in A .$$

Since

$$\nu(A) = \int_A \frac{d\nu}{d\gamma} d\gamma = \int_A \frac{d\nu}{d\lambda} d\lambda \quad \text{for all} \quad A \in A \text{ , it follows}$$

from Lemma 5 that

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\gamma}\frac{d\gamma}{d\lambda} \qquad \text{a.e.} \quad (\lambda)$$

Theorem 9

Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and assume μ is a σ-finite measure such that $\mu >> P_\theta \quad \forall \theta \in \Theta$

Then a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is sufficient if there exists a non-negative $\mathcal{A}$ measurable function s and a set $\{g_\theta, \theta \in \Theta\}$ of non-negative $\mathcal{B}$ measurable functions such that

$$\frac{dP_\theta}{d\mu} = s g_\theta \quad \forall \theta \in \Theta \quad .$$

Proof

Assume the condition of the theorem holds. First we note that we may without loss of generality assume $s \equiv 1$.

To see this, introduce the measure $\tilde{\mu}$ by $\tilde{\mu}(A) = \int_A s\,d\mu$. Then $\tilde{\mu}$ is σ-finite.

For by decomposing $\mathcal{X}$ into sets A_n of finite measure (μ) we have

$$\int_{\mathcal{X}} s\,d\mu = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} \int_{A_n \cap \{m \leq s < m+1\}} s\,d\mu$$

and every term of the double sum is finite.

Furthermore $\tilde{\mu} >> P_\theta \quad \forall \theta \in \Theta$ since $\tilde{\mu}(A) = \int_A s\,d\mu = \int_{A \cap \{s>0\}} s\,d\mu = 0$ implies

$\mu(A \cap \{s>0\}) = 0$ and hence $P_\theta(A) = \int_A g_\theta s\,d\mu = \int_{A \cap \{s>0\}} g_\theta s\,d\mu = 0 \quad \forall \theta \in \Theta$

It follows by Lemma 8 that $\dfrac{dP_\theta}{d\tilde{\mu}} = g_\theta$

We must show that $P_\theta(A|\mathcal{B}) = Y_A$ a.e. (P_θ) $\forall\theta\in\Theta$ where Y_A is a $\mathcal{B}$-measurable function for every $A \in \mathcal{A}$.

By Corollary 7 there exists a $\mathcal{B}$ measurable, non-negative function $\mu(A|\mathcal{B})$ such that

$$\int_B \mu(A|\mathcal{B})d\mu = \int_B I_A d\mu \qquad \text{for all } B \in \mathcal{B} .$$

We will prove that $P_\theta(A|\mathcal{B}) = \mu(A|\mathcal{B})$ a.e. (P_θ) $\forall\theta\in\Theta$

By the extension procedure from indicator functions to monotone limits of simple functions,

$$\int g\mu(A|\mathcal{B})d\mu = \int gI_A d\mu$$

for every non-negative $\mathcal{B}$-measurable function g .

Let $g = I_B g_\theta$. Then

$$\int_B \mu(A|\mathcal{B})g_\theta d\mu = \int_B \mu(A|\mathcal{B})dP_\theta = \int I_B g_\theta I_A d\mu =$$

$$\int_{A\cap B} g_\theta d\mu = P_\theta(A\cap B) \qquad \text{for every } B\in\mathcal{B}.$$

Thus $\mu(A|\mathcal{B}) = P_\theta(A|\mathcal{B})$ a.e. (P_θ) and the theorem is proved.

Consider now a fixed probability space $(X,\mathcal{A},P)$.

<u>Definition 10</u>

Let $\{X_t; t\in T\}$ be a family of r.v's. The r.v. Y is called <u>essential supremum</u> for the family if

(i) $X_t \leqq Y$ a.e. for all $t\in T$

(ii) $X_t \leqq Z$ a.e. for all $t\in T$ implies $Y \leqq Z$ a.e.

We write $Y = \operatorname{ess\,sup}_{t\in T} X_t$. Note that provided the essential supremum exsists, it is uniquely determined (a.e.(P)) .

If the index set T is countable, then $Y = \sup_{t\in T} X_t$ is measurable and the verification of (i) and (ii) is trivial. Thus an essential supremum always exists if our family of r.v's is countable. If T is not countable, the function $Y = \sup_{t \in T} X_t$ may not be measurable and thus not a r.v .

For example, let $X = [0,1]$, A the class of Lebesgue measurable sets on $[0,1]$ and P Lebesgue measure. Furthermore let A be a subset of $[0,1]$ such that A is not Lebesgue measurable.

Define

$$X_t(x) = \begin{cases} 1 & t = x \\ 0 & t \neq x \end{cases}$$

Then for each $t \in A$, X_t is a random variable but $I_A(x) = \sup_{t \in A} X_t$ is certainly not measurable.

The following theorem states, however, that an essential supremum still exists.

<u>Theorem 11</u>

To each family $\{X_t, t \in T\}$ of r.v's there exists an essential supremum.

<u>Proof</u>

We can without loss of generality assume $0 \leqq X_t \leqq 1$.

To see this let φ be a $1-1$ mapping of $[-\infty,\infty]$ onto $[0,1]$ (φ may be taken as $\varphi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$)

Let $\tilde{X}_t = \varphi(X_t)$. Assume we have shown the theorem for $0 \leqq X_t \leqq 1$.

Then if $\tilde{Y} = \text{ess sup } \tilde{X}_t$, $Y = \varphi^{-1}(\tilde{Y}) = \text{ess sup } X_t$.

Let

$$m(S) = E \sup_{t \in S} X_t$$

where $S \subseteq T$ and S countable and

$$m = \sup_{S: S \subseteq T, S \text{ countable}} m(S)$$

Furthermore let $\{S_n\}$ be a sequence of sets such that $S_n \subseteq T$, S countable and $m(S_n) \uparrow m$ as $n \to \infty$. If we define $S_0 = \bigcup_{n=1}^{\infty} S_n$, then $S_0 \subseteq T$ and S_0 is countable, and it follows that

$$m(S_0) \geqq m(S_n) \ \forall n . \quad \text{Letting } n \to \infty$$

$$m(S_0) \geqq m$$

Hence $m(S_0) = m$ since we by definition of m have $m \geqq m(S_0)$.

Let $Y = \sup_{t \in S_0} X_t$. We shall prove that $Y = \operatorname{ess\,sup}_{t \in T} X_t$.

(i) Let $t_1 \in T$, then obviously $Y \leqq \sup_{t \in S_0 \cup \{t_1\}} X_t$.

On the other hand $EY = m(S_0) = m = E \sup_{t \in S_0 \cup \{t_1\}} X_t$.

Thus $E(\sup_{t \in S_0 \cup \{t_1\}} X_t - Y) = 0$. Since the expression in the brackets is always non-negative, we have $\sup_{t \in S_0 \cup \{t_1\}} X_t = Y$ a.e. which implies $X_{t_1} \leqq Y$ a.e. and (i) is proved.

(ii) Assume $Z \geqq X_t$ a.e. $\forall t$. Then obviously $Z \geqq \sup_{t \in S_0} X_t = Y$ a.e.

<u>Remark</u>

It follows from the proof that there always exists a $S_0 \subseteq T$, S_0 countable such that $\operatorname{ess\,sup}_{t \in T} X_t = \sup_{t \in S_0} X_t$

Theorem 12

Let $(X, A, P_\theta, \theta \in \Theta)$ be an experiment. Assume there exists a σ-finite measure μ on A such that $P_\theta \ll \mu$ for all $\theta \in \Theta$. Then we can find a countable subset Θ_0 of Θ such that

$$P_\theta(A) = 0 \ \forall \theta \in \Theta_0 \Rightarrow P_\theta(A) = 0 \ \forall \theta \in \Theta$$

Proof

By Lemma 4, μ may be assumed to be a prob. measure.

Let $f_\theta = \frac{dP_\theta}{d\mu}$ and Θ_θ a countable subset of Θ such that

$\operatorname{ess\,sup}_\Theta f_\theta = \sup_{\theta \in \Theta_0} f_\theta$ (this is possible by the remark to Theorem 11).

Assume $P_\theta(A) = 0 \ \forall \theta \in \Theta$. Now since $0 = P_\theta(A) = \int_A f_\theta d\mu = \int f_\theta I_A d\mu$

it follows that $f_\theta I_A = 0$ a.e. (μ) $\forall \theta \in \Theta_0$. Hence $0 = I_A \sup_{\theta \in \Theta_0} f_\theta =$

$= I_A \operatorname{ess\,sup}_\theta f_\theta \geq I_A f_\theta$ a.e. (μ) $\forall \theta \in \Theta$. This implies that

$I_A f_\theta = 0$ a.e. (μ) and hence $P_\theta(A) = 0 \ \forall \theta \in \Theta$. Q.E.D.

Let $E = (X, A, P_\theta; \theta \in \Theta)$ be an experiment. E is said to be dominated if there is a σ-finite measure μ on A such that $P_\theta \ll \mu$ for all $\theta \in \Theta$. E is said to be homogeneous if

$P_{\theta_1} \sim P_{\theta_2}$ $\forall \theta_1, \theta_2$.

Theorem 13

Let $E = (X, A, P_\theta, \theta \in \Theta)$ be an experiment. Assume there exists a σ-finite measure μ on A such that $P_\theta \ll \mu$ for all $\theta \in \Theta$. Then E is dominated by a probability measure π given by $\pi = \sum_\theta \lambda_\theta P_\theta$ where $\lambda_\theta \geq 0$ for all $\theta \in \Theta$, the set of θ's for which $\lambda_\theta > 0$ is countable and $\sum_\theta \lambda_\theta = 1$

Proof

Choose a countable subset $\Theta_0 \subseteq \Theta$ with the property given in Theorem 12. Let $\lambda_\theta = 0$ for $\theta \in \Theta - \Theta_0$ and $\lambda_\theta > 0$ for $\theta \in \Theta_0$. such that $\sum_\theta \lambda_\theta = 1$. Define π by $\pi = \sum_\theta \lambda_\theta P_\theta$. Then π is a probability measure and clearly $\pi(A) = 0 \Rightarrow P_\theta(A) = 0$ for all $\theta \in \Theta_0$ which again by the choice of Θ_0 implies $P_\theta(A) = 0 \ \forall \theta \in \Theta$. Hence $P_\theta << \pi$ for all $\theta \in \Theta$.

Remark

If $\mathcal{E}$ is homogeneous, then we may take π to be one of the P_θ's , say P_{θ_0} and $\Theta_0 = \{\theta_0\}$

Theorem 14

Let $\mathcal{E} = (X, \mathcal{A}, P_\theta, \theta \in \Theta)$ be a dominated experiment. Then a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is sufficient if and only if $\frac{dP_\theta}{d\pi}$ may for each $\theta \in \Theta$ be specified $\mathcal{B}$-measurable (π is given in Theorem 13).

Proof

Assume $\frac{dP_\theta}{d\pi}$ may be specified $\mathcal{B}$-measurable for each $\theta \in \Theta$. Then by Theorem 9, $\mathcal{B}$ is sufficient. Assume $\mathcal{B}$ is sufficient. Let $A \in \mathcal{A}$, $P_\theta(A|\mathcal{B}) = Z_A$ where $0 \leq Z_A \leq 1$.
It follows that

$$\int_B Z_A d\pi = \int_B Z_A d(\Sigma\lambda_\theta P_\theta) = \sum_\theta \lambda_\theta \int_B Z_A dP_\theta = \sum_\theta \lambda_\theta P_\theta(A \cap B) = \pi(A \cap B) \quad ,$$

thus $Z_A = \pi(A|\mathcal{B})$.

Let $\frac{dP_\theta}{d\pi} = f_\theta$, $E_\pi(f_\theta|\mathcal{B}) = \tilde{f}_\theta$.

Then $\frac{dP_\theta}{d\pi} = \tilde{f}_\theta$ and hence $\frac{dP_\theta}{d\pi}$ may be specified $\mathcal{B}$-measurable since $\tilde{f}_\theta$ is $\mathcal{B}$-measurable.

The assertion holds since by definition 1 and the properties of conditional expectations:

$$\int_A \tilde{f}_\theta d\pi = \int_A E_\pi(f_\theta|\mathcal{B})d\pi = \int I_A E_\pi(f_\theta|\mathcal{B})d\pi$$

$$= \int \pi(A|\mathcal{B})E_\pi(f_\theta|\mathcal{B})d\pi = \int \pi(A|\mathcal{B})f_\theta d\pi$$

$$= \int P_\theta(A|\mathcal{B})dP_\theta = P_\theta(A)$$

Factorization Theorem 15

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},P_\theta;\theta \in \Theta)$ be an experiment and assume μ is a σ-finite measure such that $\mu \gg P_\theta$ for all $\theta \in \Theta$.
Then a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is sufficient if and only if there exists a non-negative $\mathcal{A}$-measurable function s and a set $\{g_\theta,\theta\in\Theta\}$ of non-negative $\mathcal{B}$-measurable functions such that

$$\frac{dP_\theta}{d\mu} = sg_\theta \qquad \text{for all } \theta \in \Theta .$$

Proof

The if statement is just Theorem 9. Suppose $\mathcal{B}$ is sufficient. Clearly $\mu \gg \pi$, where π is given in Theorem 13. Thus by the chain rule of Radon Nikodym derivates (Lemma 8),

$$\frac{dP_\theta}{d\mu} = \frac{d\pi}{d\mu}\frac{dP_\theta}{d\pi} \qquad \text{for all } \theta \in \Theta .$$

The theorem follows now since by Theorem 14, $dP_{\theta/d\pi}$;$\theta \in \Theta$ may all be specified $\mathcal{B}$-measurable.

Definition 16

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},P_\theta;\theta\in\Theta)$ be an experiment and let $\mathcal{B}_1$ and $\mathcal{B}_2$ be sub-σ-algebras of $\mathcal{A}$. We define an ordering $\leq$ by $\mathcal{B}_1 \leq \mathcal{B}_2 \Leftrightarrow$ For all $B_1 \in \mathcal{B}_1$ there exists a $B_2 \in \mathcal{B}_2$ such that

$$E_\theta | I_{B_1} - I_{B_2} | = E_\theta I_{B_1 \Delta B_2} = P_\theta(B_1 \Delta B_2) = 0 \qquad \text{for all } \theta \in \Theta$$

(Δ means symmetric difference) or equivalently:

For each $\mathcal{B}_1$-measurable bounded (or non-negative) function f_1 there is a $\mathcal{B}_2$-measurable bounded (or non-negative) function f_2 such that $E_\theta|f_1 - f_2| = 0$ for all $\theta \in \Theta$

i.e. $f_1 = f_2$ a.e. (P_θ) for all $\theta \in \Theta$.

Note that if $\mathcal{B}_1 \subset \mathcal{B}_2$, then $\mathcal{B}_1 \leqq \mathcal{B}_2$.

If $\mathcal{B}_1 \leqq \mathcal{B}_2$ and $\mathcal{B}_2 \leqq \mathcal{B}_1$ we say that $\mathcal{B}_1$ and $\mathcal{B}_2$ are equivalent and write $\mathcal{B}_1 \sim \mathcal{B}_2$.

A σ-algebra $\mathcal{B}_0$ is said to be <u>minimal sufficient</u> for E if $\mathcal{B}_0$ is sufficient for E and $\mathcal{B}_0 \leqq \mathcal{B}$ for all sufficient σ-algebras $\mathcal{B}$.

<u>Theorem 17</u>

Assume E is dominated and let $\mathcal{B}_0$ be the smallest σ-algebra such that the functions $f_\theta = \frac{dP_\theta}{d\pi}$ are measurable for all $\theta \in \Theta$. [π is given in Theorem 13] , i.e. $\mathcal{B}_0 = \sigma(f_\theta, \theta\in\Theta)$.

Then $\mathcal{B}_0$ is minimal sufficient for E .

<u>Proof</u>

$\mathcal{B}_0$ is sufficient by Theorem 14.

Assume $\mathcal{B}$ is sufficient for E and let $\{\tilde{f}_\theta, \theta\in\Theta\}$ be $\mathcal{B}$-measurable versions of $\frac{dP_\theta}{d\pi}$.

We shall prove that $\mathcal{B}_0 \leqq \mathcal{B}$.

It follows from Radon Nikodym's Theorem that

(1.3) $f_\theta = \tilde{f}_\theta$ a.e. (π)

By definition, $\mathcal{B}_0$ is the smallest σ-algebra containing all sets of the form $A_\theta(r) = \{x : f_\theta(x) < r\}$ for some $r \in R$ and $\theta \in \Theta$.

Define $B_\theta(r) = \{x : \tilde{f}_\theta(x) < r\}$ for $r \in R$, $\theta \in \Theta$.

Then $B_\theta(r) \in \mathcal{B}$ and by (1.3) $\pi(A_\theta(r) \Delta B_\theta(r)) = 0$ for all r, θ .

It is easy to verify that the family of sets $B_0 \in \mathcal{B}_0$ such that there exist $B \in \mathcal{B}$ with $\pi(B_0 \Delta B) = 0$ is a σ-algebra, i.e.
$\mathcal{B}' = \{B_0 \in \mathcal{B}_0 : \pi(B_0 \Delta B) = E_\pi |I_{B_0} - I_B| = 0$ for some $B \in \mathcal{B}\}$ is a σ-algebra.

Since this σ-algebra contains the sets $A_\theta(r)$, it is equal to $\mathcal{B}_0$. Hence $\mathcal{B}_0 \leq \mathcal{B}$.

Definition 18

Let $\mathcal{E}$ be an experiment.

A sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is said to be boundedly complete if for all bounded $\mathcal{B}$-measurable functions g :

$$E_\theta g = 0 \quad \text{for all} \quad \theta \in \Theta \Rightarrow g = 0 \quad \text{a.e. } (P_\theta) \quad \text{for all} \quad \theta \in \Theta .$$

Theorem 19

Let $\mathcal{E}$ be an experiment.

Assume that $\mathcal{B}$ is sufficient and boundedly complete.

If $\mathcal{C}$ is sufficient and $\mathcal{C} \leq \mathcal{B}$, then $\mathcal{B} \sim \mathcal{C}$.

Remark

Any sufficient σ-algebra $\mathcal{B}$ such that $\mathcal{C} \sim \mathcal{B}$ whenever $\mathcal{C}$ is a sufficient sub-σ-algebra of $\mathcal{B}$ is actually minimal sufficient. See Burkholder (1961). We will prove this here (Corollary 20) only when $\mathcal{B}$ is boundedly complete.

Proof

It sufficies to prove that $\mathcal{B} \leq \mathcal{C}$.

Let $B \in \mathcal{B}$. By definition 2 there exists a $\mathcal{C}$-measurable Y such that

$$P_\theta(B|\mathcal{C}) = Y \quad \text{a.e. } (P_\theta) \quad \text{for all} \quad \theta \in \Theta.$$

Let

$$C = \{x : Y(x) = 1\} . \quad \text{Clearly} \quad C \in \mathcal{C} .$$

Since $\mathcal{C} \leqq \mathcal{B}$ there is a $\mathcal{B}$-measurable function Z such that $Z = Y$ a.e. (P_θ) ; $\theta \in \Theta$. Z is bounded (a.e.) since Y is . Furthermore, for any $\theta \in \Theta$

$$\int Z dP_\theta = \int Y dP_\theta = \int P_\theta(B|\mathcal{C}) dP_\theta = P_\theta(B)$$

Hence $\int (I_B - Z) dP_\theta = 0$ for all θ .

Since $\mathcal{B}$ is boundedly complete it follows that

$$I_B = Z \quad \text{a.e.} \quad (P_\theta) \quad \text{for all} \quad \theta .$$

Hence $Y = I_B$ a.e. (P_θ) for all θ so that

$$P_\theta(B \Delta C) = 0 \quad \text{for all} \quad \theta .$$

<u>Corollary 20</u>

Let $\mathcal{E}$ be a dominated experiment.

If $\mathcal{B}$ is sufficient and boundedly complete, then $\mathcal{B}$ is minimal sufficient.

<u>Proof</u>

Let $\mathcal{B}_0$ be given as in Theorem 17. Then $\mathcal{B}_0 \leqq \mathcal{B}$. The corollary follows from Theorem 19.

We will now consider sufficiency in terms of operational characteristics. First we give some basic definitions (see Torgersen & Lindquist (1975)).

Definition 21

Let $(\mathcal{X},\mathcal{A})$, $(\mathcal{Y},\mathcal{B})$ be measurable spaces. A Markov-kernel from $(\mathcal{X},\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$ is a function ρ from $\mathcal{B}\times\mathcal{X}$ to R such that

(i) for each $x\in\mathcal{X}$, $\rho(\cdot|x)$ is a probability measure on $\mathcal{B}$

(ii) for each $B\in\mathcal{B}$, $\rho(B|\cdot)$ is a (bounded) measurable function on $(\mathcal{X},\mathcal{A})$

Definition 22

A decision space is a measurable space $(T,\mathcal{S})$. The elements of T are called decisions.

Definition 23

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},P_\theta:\theta\in\Theta)$ be an experiment. A decision rule ρ is a Markov-kernel from $\mathcal{E}$ to $(T,\mathcal{S})$.

The operational characteristic of ρ is the function

$$OC_\rho : \mathcal{S}\times\Theta\to R \quad \text{given by}$$

$$OC_\rho(S|\theta) = \int\rho(S|x)P_\theta(dx) = E_\theta\rho(S|\cdot)$$

Definition 24

Let $(\mathcal{X},\mathcal{A},P)$ be a probability space and $\mathcal{B}$ a sub-σ-algebra of $\mathcal{A}$. Then a Markov-kernel π from $(\mathcal{X},\mathcal{B})$ to $(\mathcal{X},\mathcal{A})$ is called a regular conditional probability for $\mathcal{A}$ given $\mathcal{B}$ provided $\pi(A|\cdot)$ is a version of $P(A|\mathcal{B})$ for each $A\in\mathcal{A}$

Note that if π is a regular conditional probability for $\mathcal{A}$ given $\mathcal{B}$, then $\int X(x')\pi(dx'|\cdot)$ is a version of $E(X|\mathcal{B})$ for each X such that EX exists. This may be seen by the standard extension procedure from indicator functions to simple functions and monotone limits of simple functions.

Theorem 25

Let $\mathcal{E} = (X, \mathcal{A}, P_\theta, \theta \in \Theta)$ be an experiment and let $\mathcal{B}$ be a **sufficient** sub-σ-algebra of $\mathcal{A}$. Suppose π is a regular conditional probability for $\mathcal{A}$ given $\mathcal{B}$ for all $\theta \in \Theta$. Let $(\mathcal{T}, \mathcal{S})$ be any decision space and ρ any decision rule from $\mathcal{E}$ to $(\mathcal{T}, \mathcal{S})$. Then the $\mathcal{B}$-measurable decision rule $\rho \circ \pi$ defined by

$$(\rho \circ \pi)(S|\cdot) = \int \rho(S|x')\pi(dx'|\cdot)$$

has the same operating characteristic as ρ.

Remark

By Theorem 26 there always exists a regular conditional probability π for $\mathcal{A}$ given $\mathcal{B}$ for all $\theta \in \Theta$ if $(X, \mathcal{A})$ is Euclidean and $\mathcal{B}$ is sufficient. $(X, \mathcal{A})$ is called Euclidean if either X is enumerable, with $\mathcal{A}$ being the class of all subsets or $(X, \mathcal{A})$ is Borel-isomorphic to the real line (i.e. there exists a $\mathcal{A}$-measurable function $\psi : X \to R$ such that ψ is one-to-one and onto, and ψ^{-1} is measurable).

It is known (Parthasarathy) that $(X, \mathcal{A})$ is Euclidean whenever X is a Borel-subset of a complete separable metric space and $\mathcal{A}$ is the class of Borel subsets of X. (The Borel subsets of a metric space is the σ-algebra of sets generated by the open sets).

Proof

Since $(\rho \circ \pi)(S|\cdot)$ is a version of $E_\theta(\rho(S|\cdot)|\mathcal{B})$ for all $S \in \mathcal{S}$, it follows that

$$OC_{\rho\circ\pi}(S|\theta) = \int (\rho\circ\pi)(S|\cdot)dP_\theta = \int \rho(S|\cdot)dP_\theta = OC_\rho(S|\theta)$$

Reffering to the set up in Theorem 25 we may conclude that there to any decision rule corresponds a $\mathcal{B}$-measurable decision rule with the some operating characteristic and in particular with the some risk function (provided a loss is defined). It follows that the experiment $\mathcal{E}$ may be recovered from the restriction $\mathcal{E}|\mathcal{B}$ by performing a randomization according to the known chance mechanisme π .

If we are primarily interested in a particular decision space $(\mathcal{T},\mathcal{S})$, then we may avoid any assumption on the existence of regular conditional probabilities in $\mathcal{E}$ provided $(\mathcal{T},\mathcal{S})$ is Euclidean.

Theorem 26

Let the σ-algebra $\mathcal{B}$ be sufficient in the experiment $\mathcal{E} = (X,\mathcal{A},P_\theta,\theta\in\Theta)$ and consider a decision rule ρ from $\mathcal{E}$ to the Euclidean decision space $(\mathcal{T},\mathcal{S})$. Then the conditional expectations $E_\theta(\rho(S|\cdot)|\mathcal{B})$ may be regularized to a decision rule $\tilde{\rho}(S|\cdot) = E_\theta(\rho(S|\cdot)|\mathcal{B})$ having the same operating characteristic as ρ .

Proof

We may without loss of generality assume that $\mathcal{T}=$ the real line and that $\mathcal{S}$ is the class of Borel subsets of $\mathcal{T}$.

Choose, for each rational number $r \in Q =$ the set of rational numbers a version $\tilde{\rho}((-\infty,r]|\cdot)$ of $E_\theta(\rho((-\infty,r]|\cdot)|\mathcal{B})$ We may also, after a possible redefinitions of $\tilde{\rho}$ on a null set, assume that the following hold for all x:

(i) $0 \leqq \tilde{\rho}((-\infty,r]|x) \leqq \tilde{\rho}((-\infty,s]|x) \leqq 1$ when $r,s \in Q$ and $r \leqq s$

(ii) $\tilde{\rho}((-\infty,r+\frac{1}{n}|x) \downarrow \tilde{\rho}((-\infty,r]|x)$ as $n \to \infty$

(iii) $\tilde{\rho}((-\infty,r]|x) \to 1$ as $r \to \infty$

(iv) $\tilde{\rho}((-\infty,r]|x) \to 0$ as $r \to -\infty$

Put for each real number t : $\tilde{\rho}((-\infty,t]\,|\,\cdot) = \inf\{\tilde{\rho}((-\infty,r]\,|\,\cdot) : r \geq t\}$ (There is clearly no conflict here when $t \in Q$).

Clearly $\tilde{\rho}((-\infty,t]\,|\,\cdot)$ is a version of $E_\theta(\rho((-\infty,t]\,|\,\cdot)\,|\,\mathcal{B})$ for each $t \in \mathcal{T}$.

Let finally for each fixed x $\tilde{\rho}(S|x)$ be the measure assigned to to S by the probability measure on $\mathcal{S}$ determined by the distribution function $t \to \tilde{\rho}((-\infty,t]\,|\,x)$. As the class of sets S such that $\tilde{\rho}(S|\cdot)$ is a version of $E_\theta(\rho(S|\cdot)|\mathcal{B})$ is clearly a λ-system and contains the π-system of intervals $\{(-\infty,t], t\in\mathcal{T}\}$, we find that $\tilde{\rho}(S|\cdot)$ is a version of $E_\theta(\rho(S|\cdot)|\mathcal{B})$ for all $S \in \mathcal{S}$.

Hence

$$OC_{\tilde{\rho}}(S|\theta) = \int \tilde{\rho}(S|\cdot)dP_\theta = \int \rho(S|\cdot)dP_\theta = OC_\rho(S|\theta) \quad \text{for all} \quad S \in \mathcal{S}$$

and all $\theta \in \Theta$.

References.

Bahadur, R. R. (1954) Sufficiency and statistical decision functions. Ann.Math. Statist, 25, 423-462

Bahadur, R. R. (1955) A characterization of sufficiency. Ann. Math. Statist, 26, 286-293

Burkholder, B.L. (1961) Sufficiency in the randomized case. Ann. Math. Statist, 32, 1191-1200

Dynkin, E. B. (1951) Necessary and sufficient statistics for a family of probability distributions (in Russian). Uspetri Mat. Nauk. (N.S.) 6 no 1, 41, 68-90. English translation: Selected translations in Mathematical Society, Providence, Rhode Island, 1961.

Halmos, P. and Savage, L.J. (1949), Application of the Radon-Nikodym theorem to the theory of sufficient statistics. Ann. Math. Statist., 20, 225-241.

Heyer, H. (1973), Mathematische Theorie Statistischer Experimente. Springer Verlag, Berlin.

Landers, D. (1974), Minimal sufficient statistics for families of product measure. Ann. Statist., 2, 1335-1339.

Le Cam, L. (1964), Sufficiency and approximate sufficiency. Ann. Math. Statist., 35, 1419-1455.

Neveu, J. C. (1965), Mathematical foundations of the calcus of probability. Holden Day.

Parthasarathy, K.R. (1967), Probability measures on metric spaces. Academic Press.

Siebert, E. (1979), Statistical experiments and their conical measures. Z. Wahrscheinlichkeits theorie, verw. Gebiete 46, 247-258.

Torgersen, E.N. & Lindquist, B. (1975), Notes on comparison of statistical experiments. Statistical Memoirs. University of Oslo.

Torgersen, E.N. (1976), Comparison of Statistical experiments. Scand. Journ. of Statist., 3, 186-208.

Wald, A. (1950), Statistical decision functions. Wiley, New York.