The Annals of Probability
1974, Vol. 2, No. 3, 490–494

CONVEXITY AND CONDITIONAL EXPECTATIONS

BY J. PFANZAGL

University of Cologne

If a n-dimensional function is with probability one in a convex set, the same holds true for the conditional expectation (with respect to any sub-σ-field). An extreme point of this convex set can be assumed by the conditional expectation only if it is assumed by the original function and if this function is partially measurable with respect to the conditioning sub-σ-field.

These results are used to prove Jensen's inequality for conditional expectations of n-dimensional functions, and to give a condition for strict inequality.

NOTATIONS. For any set A let 1_A denote its indicator function and A^c its complement. For any measure μ, let μ^* and μ_* denote the pertaining outer and inner measure, respectively. $(X, \mathscr{A}, P)$ denotes a probability space. For any sub-σ-field $\mathscr{A}_0 \subset \mathscr{A}$, the symbol $P^{\mathscr{A}_0}f$ denotes the class of all conditional expectations of f, given $\mathscr{A}_0$. Instead of $P^{\mathscr{A}_0}1_A$ we shall write $P^{\mathscr{A}_0}(A)$. Let $\mathbb{R}^n$ denote the n-dimensional Euclidean space and $\mathscr{B}^n$ its Borel field. For $n = 1$ we write $\mathbb{R}$ and $\mathscr{B}$ instead of $\mathbb{R}^1$, $\mathscr{B}^1$. The symbol π_i denotes the projection from $\mathbb{R}^n$ into its ith component. For any convex set $C \subset \mathbb{R}^n$, C_e denotes the set of extreme points of C. The measure induced by $f: X \to \mathbb{R}^n$ and $P \mid \mathscr{A}$ is $B \to P(f^{-1}B)$, $B \in \mathscr{B}^n$. A function $f: X \to \mathbb{R}$ is P-integrable if its integral, $P(f)$, is finite.

THEOREM 1. *Let $f_i: X \to \mathbb{R}$, $i = 1, \cdots, n$, be P-integrable functions and $C \in \mathscr{B}^n$ a convex set for which*

$$P\{x \in X: (f_1(x), \cdots, f_n(x)) \in C\} = 1 .$$

Then for arbitrary versions $g_i \in P^{\mathscr{A}_0}f_i$, $i = 1, \cdots, n$,

(i) $P\{x \in X: (g_1(x), \cdots, g_n(x)) \in C\} = 1$;

(ii) *f_i and g_i coincide P-a.e. on the set $\{x \in X: (g_1(x), \cdots, g_n(x)) \in C_e\}$.*

For $\mathscr{A}_0 = \{\emptyset, X\}$, Theorem 1 reduces to the corresponding theorem for integrals. For this special case, part (i) is well known and will, in fact, be needed as the following lemma in the proof of Theorem 1.

LEMMA. *Let $C \subset \mathbb{R}^n$ be a convex set (not necessarily measurable) and $Q \mid \mathscr{B}^n$ a probability measure such that $Q^*(C) = 1$. Then $(Q(\pi_1), \cdots, Q(\pi_n))$, the barycenter of Q, belongs to C.*

PROOF. See Ferguson (1967) page 74, Lemma 3.

Received June 4, 1973; revised September 11, 1973.

AMS 1970 *subject classifications.* Primary 62B99; Secondary 52A40.

Key words and phrases. Conditional expectations, convex sets.

The question suggests itself whether the detour via this lemma can be avoided, i.e. whether a direct proof of Theorem 1 is feasible. Such a proof can, in fact, be easily given if the convex set C is closed (and therefore the intersection of closed hyperplanes). It is, however, rather difficult to obtain a direct proof for measurable convex sets C in general.

Part (ii) of Theorem 1 is well known for integrals and compact convex sets C (see e.g. Bourbaki (1965) page 218, Corollaire). The interpretation of part (ii) is that positive probability induced by the map $x \to (g_1(x), \cdots, g_n(x))$ in extreme points of C is a remainder of the probability induced in these points by $x \to (f_1(x), \cdots, f_n(x))$, which is not removed because the map $x \to (f_1(x), \cdots, f_n(x))$ is $\mathscr{A}_0$-measurable on $\{x \in X: (g_1(x), \cdots, g_n(x)) \in C_e\}$. The interest in (ii) is motivated by its relationship to strict inequality in Jensen's inequality.

Proof of Theorem 1. (i) Let $f \equiv (f_1, \cdots, f_n): X \to \mathbb{R}^n$. According to Doob (1952) page 29, Theorem 9.4, there exists a function $q: \mathscr{B}^n \times X \to [0, 1]$ such that

$$q(\cdot, x) \mid \mathscr{B}^n \quad \text{is a probability measure for each} \quad x \in X$$
$$q(B, \cdot) \in P^{\mathscr{A}_0}(f^{-1}B) \qquad \text{for each} \quad B \in \mathscr{B}^n .$$

For each Borel-measurable function $h: \mathbb{R}^n \to \mathbb{R}$ for which $h \circ f$ is P-integrable,

$$q(h, \cdot) \in P^{\mathscr{A}_0}(h \circ f) .$$

Hence

$$q(\pi_i, \cdot) \in P^{\mathscr{A}_0} f_i \qquad \text{for } i = 1, \cdots, n .$$

Since $P(f^{-1}C) = 1$ and $q(C, \cdot) \in P^{\mathscr{A}_0}(f^{-1}C)$, there exists a P-null set $N \in \mathscr{A}_0$ such that $q(C, x) = 1$ for all $x \notin N$. As $q(\cdot, x) \mid \mathscr{B}^n$ is a probability measure, the lemma implies for each $x \notin N$:

$$(q(\pi_1, x), \cdots, q(\pi_n, x)) \in C .$$

Hence the assertion holds true for the particular versions $q(\pi_i, \cdot) \in P^{\mathscr{A}_0} f_i$ and therefore for arbitrary versions $g_i \in P^{\mathscr{A}_0} f_i$.

(ii) It seems to be unknown whether $C \in \mathscr{B}^n$ implies $C_e \in \mathscr{B}^n$. However, $C - C_e = \{\frac{1}{2}r + \frac{1}{2}s : r, s \in C, r \neq s\}$, i.e. $C - C_e$ is the image of the measurable set $\{(r, s): r, s \in C, r \neq s\} \subset \mathbb{R}^{2n}$ under the continuous map $(r, s) \to \frac{1}{2}r + \frac{1}{2}s$ and hence analytic (see Bressler and Sion, page 216, Theorem 5.13). This implies that $C - C_e$ and hence also C_e belongs to the completion of $\mathscr{B}^n$ with respect to any finite measure on $\mathscr{B}^n$ (see Bressler and Sion, page 226, Theorem 6.9). Hence the set C_e belongs to the completion of $\mathscr{B}^n$ with respect to the measure induced by P and g. Since g is $\mathscr{A}_0$-measurable, $A_e \equiv g^{-1}C_e$ belongs to the completion of $\mathscr{A}_0$ with respect to P.

We shall show that $g_i 1_{A_e} = f_i 1_{A_e}$ P-a.e. for $i = 1, \cdots, n$. To this end it suffices to prove that

$$P(1_{A_e \cap A} g_i) = P(1_{A_e \cap A} f_i) \qquad \text{for all} \quad A \in \mathscr{A}, i = 1, \cdots, n .$$

Let $A \in \mathscr{A}$ be arbitrary. If $P(A) = 0$ or $P(A) = 1$, this is obviously true. Hence assume $P(A) \in (0, 1)$, and choose $c_A \in P^{\mathscr{A}_0} 1_A$ such that $c_A(x) \in [0, 1]$, $x \in X$. Let probability measures $P' | \mathscr{A}$ and $P'' | \mathscr{A}$ be defined by

$$P'(D) \equiv P(D \cap A)/P(A) \text{ resp. } P''(D) \equiv P(D \cap A^c)/P(A^c)\,, \qquad D \in \mathscr{A}\,.$$

Let $g_i' \in P'^{\mathscr{A}_0} f_i$, $g_i'' \in P''^{\mathscr{A}_0} f_i$ and

$$\hat{g}_i \equiv c_A g_i' + (1 - c_A) g_i''\,, \qquad i = 1, \cdots, n\,.$$

It is easy to check that $\hat{g}_i \in P^{\mathscr{A}_0} f_i$. Hence there exists $A_* \in \mathscr{A}_0$ with $A_* \subset A_e$ and $P(A_e - A_*) = 0$ such that $x \in A_*$ implies $(\hat{g}_1(x), \cdots, \hat{g}_n(x)) \in C_e$.

Let $D_0 \equiv \{x \in X : c_A(x) = 0\}$, $D_1 \equiv \{x \in X : c_A(x) = 1\}$ and $D_* \equiv D_0^c \cap D_1^c$. For $x \in A_* \cap D_*$ we have $(\hat{g}_1(x), \cdots, \hat{g}_n(x)) \in C_e$ and $0 < c_A(x) < 1$, hence $g_1'(x) = g_i''(x)$ for $i = 1, \cdots, n$. Furthermore, $P(D_0 \cap A) = P(1_{D_0} c_A) = 0$ and $1_{D_1}(x)(1 - c_A(x)) = 0$, $x \in X$. Hence

$$\begin{aligned} P(1_{A_* \cap A} \hat{g}_i) &= P(1_{A_* \cap A} c_A g_i') + P(1_{A_* \cap A}(1 - c_A) g_i'') \\ &= P(1_{A_* \cap A \cap D_1} g_i') + P(1_{A_* \cap A \cap D_*} c_A g_i') + P(1_{A_* \cap A \cap D_*}(1 - c_A) g_i'') \\ &= P(1_{A_* \cap A \cap D_1} g_i') + P(1_{A_* \cap A \cap D_*} g_i') = P(1_{A_* \cap A} g_i') \\ &= P'(1_{A_*} g_i') P(A) = P'(1_{A_*} f_i) P(A) = P(1_{A_* \cap A} f_i)\,. \end{aligned}$$

Since $1_{A_*} = 1_{A_e}$ P-a.e. and $\hat{g}_i = g_i$ P-a.e., this implies

$$P(1_{A_e \cap A} g_i) = P(1_{A_e \cap A} f_i)\,, \qquad \text{for } i = 1, \cdots, n\,. \;\square$$

In the particular case of integrals (i.e. for $\mathscr{A}_0 = \{\varnothing, X\}$) Theorem 1 holds true for arbitrary (i.e. not necessarily measurable) convex sets C. Therefore it seems worthwhile to mention that in the general case of conditional expectations the measurability assumption for C cannot be dispensed with. If the convex set is not measurable it may occur that

$$\{x \in X : (f_1(x), \cdots, f_n(x)) \in C\} = X\,, \qquad \text{but}$$
$$\{x \in X : (g_1(x), \cdots, g_n(x)) \in C\} = \varnothing\,.$$

EXAMPLE. Let $X = [0, 1)$ and let $M \subset [0, 1)$ be a set with $\lambda^*(M) = 1$ and $\lambda_*(M) = 0$, where λ denotes the Lebesgue-measure on $[0, 1)$ (for existence see Halmos (1950) page 70, Theorem E). Let $\mathscr{A}_0 \equiv \mathscr{B} \cap [0, 1)$ and $\mathscr{A} \equiv \{B_1 \cap M + B_2 \cap M^c : B_1, B_2 \in \mathscr{A}_0\}$. We have $\mathscr{A}_0 \subset \mathscr{A}$. Let $P | \mathscr{A}$ be defined by $P(B_1 \cap M + B_2 \cap M^c) \equiv \frac{1}{2}\lambda(B_1) + \frac{1}{2}\lambda(B_2)$, $B_1, B_2, \in \mathscr{A}_0$. It is easy to see that $P | \mathscr{A}$ is a well-defined probability measure fulfilling $P | \mathscr{A}_0 = \lambda | \mathscr{A}_0$. Let $f_1(x) = \cos 2\pi x$, $f_2(x) = \sin 2\pi x$, $f_3(x) = 1_M(x)$, $x \in X$. We remark that f_1, f_2 are $\mathscr{A}_0$-measurable, f_3 is $\mathscr{A}$-measurable. Finally let

$$\begin{aligned} C \equiv\ & \{(r_1, r_2, r_3) \in \mathbb{R}^3 : r_1^2 + r_2^2 < 1, 0 \leqq r_3 \leqq 1\} \\ & \cup \{(r_1, r_2, r_3) \in \mathbb{R}^3 : r_1 = \cos 2\pi t, r_2 = \sin 2\pi t, t \in M^c; r_3 = 0\} \\ & \cup \{(r_1, r_2, r_3) \in \mathbb{R}^3 : r_1 = \cos 2\pi t, r_2 = \sin 2\pi t, t \in M; r_3 = 1\}\,. \end{aligned}$$

The set C is convex, but not in $\mathscr{B}^3$. We have $(f_1(x), f_2(x), f_3(x)) \in C$ for all

$x \in X$. Let $g_i = f_i$, $i = 1, 2$, and $g_3 \equiv \frac{1}{2}$. Then $g_i \in P^{\mathscr{A}_0} f_i$, $i = 1, 2, 3$. Nevertheless. $(g_1(x), g_2(x), g_3(x)) \notin C$ for all $x \in X$.

THEOREM 2. (Jensen's inequality). *Let $f_i: X \to \mathbb{R}$, $i = 1, \cdots, n$, be P-integrable functions and $C \in \mathscr{B}^n$ a convex set such that $(f_1(x), \cdots, f_n(x)) \in C$ for every $x \in X$. Let, furthermore, $\varphi: C \to \mathbb{R}$ be a measurable convex function such that $\varphi^{\mathscr{A}_0}(f_1, \cdots, f_n)$ is P-integrable.*

Then for arbitrary versions $g_i \in P^{\mathscr{A}_0} f_i$, $i = 1, \cdots, n$ and $h_0 \in P^{\mathscr{A}_0}(\varphi \circ (f_1, \cdots, f_n))$,

(i) $\varphi \circ (g_1, \cdots, g_n) \leqq h_0$ *P*-a.e.

(ii) *If φ is strictly convex, then $f_i = g_i$ P-a.e. on the set $\{x \in X: \varphi(g_1(x), \cdots, g_n(x)) = h_0(x)\}$.*

For integrals rather than conditional expectations part (i) of Theorem 2 is well known (see e.g. Ferguson (1967) page 76, Lemma 1). For conditional expectations, part (i) is known for $n = 1$ (see e.g. Doob (1952) page 33). The proof given below which rests upon Theorem 1 seems particularly simple.

Part (ii) of this theorem assures that—for strictly convex functions φ—the inequality given in (i) is strict with the possible exception of a set on which all f_i are "$\mathscr{A}_0$-measurable".

PROOF OF THEOREM 2.

(i) Since φ is convex and measurable, the set

$$D \equiv \{(r_0, r_1, \cdots, r_n) \in \mathbb{R} \times C: r_0 \geqq \varphi(r_1, \cdots, r_n)\}$$

is convex and measurable. Since $(\varphi \circ (f_1, \cdots, f_n)(x), f_1(x), \cdots, f_n(x)) \in D$ for each $x \in X$, we obtain from Theorem 1 (i) that $P\{x \in X: (h_0(x), g_1(x), \cdots, g_n(x)) \in D\} = 1$ for arbitrary versions $h_0 \in P^{\mathscr{A}_0}\varphi \circ (f_1, \cdots, f_n)$ and $g_i \in P^{\mathscr{A}_0} f_i$, $i = 1, \cdots, n$.

(ii) If φ is strictly convex, $(\varphi(r_1, \cdots, r_n), r_1, \cdots, r_n)$ is an extreme point of D for all $(r_1, \cdots, r_n) \in C$. Hence $h_0(x) = \varphi(g_1(x), \cdots, g_n(x))$ implies that $(h_0(x), g_1(x), \cdots, g_n(x))$ is an extreme point of D, i.e. $A_e \equiv \{x \in X: h_0(x) = \varphi(g_1(x), \cdots, g_n(x))\} \subset \{x \in X: (h_0(x), g_1(x), \cdots, g_n(x)) \in D_e\}$, so that by Theorem 1 (ii)

$$(1_{A_e}\varphi(f_1, \cdots, f_n), 1_{A_e}f_1, \cdots, 1_{A_e}f_n) = (1_{A_e}h_0, 1_{A_e}g_1, \cdots, 1_{A_e}g_n) \quad P\text{-a.e.}$$

and therefore $(1_{A_e}f_1, \cdots, 1_{A_e}f_n) = (1_{A_e}g_1, \cdots, 1_{A_e}g_n)$ *P*-a.e.

Acknowledgment. The author wishes to thank Mr. D. Landers and Mr. L. Rogge for suggesting a remarkable simplification in the proof of Theorem 1 (i).

REFERENCES

BOURBAKI, N. (1965). *Eléments de Mathématique, Livre VI*, Intégration, 2nd ed. Hermann, Paris, Chapters 1-4.

BRESSLER, D. W. and SION, M. (1964). The current theory of analytic sets. *Canad. J. Math.* **16** 207-230.

DOOB, L. J. (1952). *Stochastic Processes.* Wiley, New York.

FERGUSON, T. S. (1967). *Mathematical Statistics. A Decision Theoretic Approach.* Academic Press, New York.
HALMOS, P. R. (1950). *Measure Theory.* Van Nostrand, Princeton.

MATHEMATISCHES INSTITUT
DER UNIVERSITAT ZU KOLN
5 KOLN 41
WEYERTAL 86–90
GERMANY