PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 20, Fasc. 2 (2000), pp. 287–291

ON DISTRIBUTIONS OF CONDITIONAL EXPECTATIONS

BY

ADAM PASZKIEWICZ* (ŁÓDŹ)

Abstract. Let F and G be distribution functions on $\boldsymbol{R}$. Then there exist a random variable X and a σ-field $\mathfrak{A}$ satisfying $P(X < a) = F(a)$, $P(E(X|\mathfrak{A}) < a) = G(a)$ iff $\int_{(a,\infty)} (F(t) - G(t))\,dt \leqslant 0 \leqslant \int_{(-\infty,a)} (F(t) - G(t))\,dt$ for any $a \in \boldsymbol{R}$. The consideration is kept on a rather elementary level.

AMS 1991 Subject Classification: 60E05.

Key words and phrases: distribution of random variable, conditional expectation.

All distributions on $\boldsymbol{R}$ used in the paper have finite first moments. We shall give an elementary proof of the following

THEOREM 1. *For any distribution functions F and G on $\boldsymbol{R}$ the following conditions are equivalent:*

(i) *there exist a random variable X and a σ-field of events $\mathfrak{A}$ satisfying*

$$P(X < a) = F(a), \quad P(E(X|\mathfrak{A}) < a) = G(a) \quad \text{for } a \in \boldsymbol{R};$$

(ii)

$$\int_{(a,\infty)} (F(x) - G(x))\,dx \leqslant 0 \leqslant \int_{(-\infty,a)} (F(x) - G(x))\,dx \quad \text{for all } a \in \boldsymbol{R}.$$

We start with some comments. Let distribution functions F and G correspond to random variables Y and Z defined on a classical probability space $\Omega = \{\omega_1, \ldots, \omega_N\}$, $P(\omega_i) = 1/N$. Then condition (ii) is equivalent to the classical majorization condition for sequences $(Y(\omega_i)) \prec (Z(\omega_i))$. In this case Theorem 1 can be obtained by a classical and old construction of a suitable bistochastic matrix. In [3], a number of other relations between majorization and matrix theory are described. Some non-expected applications are also given.

* Faculty of Mathematics, University of Łódź. Research supported by KBN grant 2 P03A 023 15.

On the other hand, condition (ii) is equivalent to the famous Karamata condition:

$$\int \phi(x)\,dF(x) \leqslant \int \phi(x)\,dG(x) \quad \text{for any convex positive function } \phi,$$

given at first at [2]. Thus, for distributions concentrated on a bounded intervals, Theorem 1 is a very special case of (for example) Theorem T2 in [4]. But the general theory is abstract and based on the Choquet theorem. So, we show perhaps as much as possible on a completely elementary level. Distributions of some systems of random variables are specially interesting. Thus the Karamata condition is still attractive for probabilists and new methods appear; see [1] and [5].

Now we establish some notation.

For convenience, random variables appearing in different formulas in the paper can be defined on different probability spaces. We use the standard notation: $p_X(A) = P(X \in A)$, $F_X(a) = P(X \leqslant a)$, $F_p(a) = p(-\infty, a)$, $A \in$ Borel $\boldsymbol{R}$, $a \in \boldsymbol{R}$, for a random variable X and a probability distribution p on $\boldsymbol{R}$.

We denote by $X^1(x^1, x^2) = x^1$ and $X^2(x^1, x^2) = x^2$ the coordinate functions on $\boldsymbol{R}^2$ and, for a distribution d on $\boldsymbol{R}^2$, by d_{X^1} and d_{X^2} the margin distributions, by $d(X^2 \mid X^1)$ and $d(X^2 \mid X^1 = t)$ the conditional distributions, and by $E_d(X^2 \mid X^1)$ and $E_d(X^2 \mid X^1 = t)$ the conditional expectations. A special role will be played by the class

$$\mathscr{S} = \{F;\ F \text{ is a simple distribution function on } \boldsymbol{R}\}; \tag{1}$$

thus $F \in \mathscr{S}$ if it describes probability concentrated on a finite set.

LEMMA 2. *If conditions* (i) *and* (ii) *are equivalent for any* $F, G \in \mathscr{S}$, *then* (i) *and* (ii) *are equivalent for any distribution functions* F *and* G *on* $\boldsymbol{R}$.

Proof. Let F and G be any distribution functions satisfying (i). Denote by d the joint distribution of the random variables $E(X \mid \mathfrak{A})$, X on $\boldsymbol{R}^2$. Let (d_n) be a sequence of distributions on $\boldsymbol{R}^2$ concentrated on finite sets such that $d_n \to d$ weakly as $n \to \infty$, and

$$E_{d_n}(X^2 \mid X^1 = t) = t \quad \text{if only} \quad d_{n,X^1}\{t\} > 0.$$

Then, for random variables X^1 and X^2 on the probability space ($\boldsymbol{R}^2$, Borel $\boldsymbol{R}^2$, d_n), the distribution functions $F^n = F_{X^2}$ and $G^n = F_{X^1} = F_{E(X^2 \mid X^1)}$ satisfy (i), and thus (ii). Moreover, $F^n \to F$ and $G^n \to G$ weakly, and condition (ii) is satisfied for the original distribution functions F and G.

Now, let the condition (ii) be satisfied. Then $\int t\,dF(t) = \int t\,dG(t)$ and we denote this value by m. For any sequences (F^n) and (G^n) in $\mathscr{S}$, satisfying

$F^n 1_{(-\infty,m)}$ decrease to $F 1_{(-\infty,m)}$,
$F^n 1_{(m,\infty)}$ increase to $F 1_{(m,\infty)}$,
$G^n 1_{(-\infty,m)}$ increase to $G 1_{(-\infty,m)}$,
$G^n 1_{(m,\infty)}$ decrease to $G 1_{(m,-\infty)}$,

we have (ii) for F^n and G^n instead of F and G, respectively. Thus there exist some distributions d_n on $\boldsymbol{R}^2$, appearing as the joint distributions of pairs $E((X_n \mid \mathfrak{A}_n), X_n)$, satisfying

$$d_{n,X^2} = F^n, \quad d_{n,X^1} = G^n, \quad E_{d_n}(X^2 \mid X^1) = X^1.$$

Obviously, the sequence (d_n) is a tide one and there exists a weak concentration point d. For the probability space ($\boldsymbol{R}^2$, Borel $\boldsymbol{R}^2$, d), the coordinates X^1 and X^2 satisfy

$$F_{X^2} = F, \quad F_{E_d(X^2 \mid \mathfrak{A})} = G \quad \text{for } \mathfrak{A} = \sigma(X^1).$$

To prove Theorem 1, it is enough to show some properties of the class $\mathscr{S}$, often elementary, concerned with conditions (i) and (ii). For $F \in \mathscr{S}$, let us put

(2) $\quad \mathscr{C}(F) = \{G \in \mathscr{S};$ (i) is satisfied$\}, \qquad \mathscr{S}(F) = \{G \in \mathscr{S};$ (ii) is satisfied$\}$.

Remark 3. If $G \in \mathscr{S}(F)$ and $H \in \mathscr{S}(G)$, then $H \in \mathscr{S}(F)$.

LEMMA 4. *For a random variable $X = \sum_{1 \leqslant i \leqslant n} \lambda_i 1_{A_i}$ with $(A_1, \ldots, A_n)$ being a partition of Ω on disjoint events, and for $\mathfrak{A} = \sigma(A_1 \cup A_2, A_3, \ldots, A_n)$ we have*

$$F_{E(X \mid \mathfrak{A})} \in \mathscr{S}(F_X).$$

Proof. An elementary calculation is sufficient. One can assume that $\lambda_1 < \lambda_2$, and check that

$$F_X - F_{E(X \mid \mathfrak{A})}(x) = \begin{cases} 0 & \text{for } x < \lambda_1, \\ P(A_1) & \text{for } \lambda_1 \leqslant x < \lambda, \\ -P(A_2) & \text{for } \lambda \leqslant x < \lambda_2, \\ 0 & \text{for } x \geqslant \lambda_2 \end{cases}$$

for $\lambda = (\lambda_1 P(A_1) + \lambda_2 P(A_2))(P(A_1) + P(A_2))^{-1}$.

LEMMA 5. *According to* (1) *and* (2), *$\mathscr{C}(F)$ is contained in $\mathscr{S}(F)$ for $F \in \mathscr{S}$.*

Proof. Let $F_X, F_{E(X \mid \mathfrak{A})} \in \mathscr{S}$ and, for simplicity, $\mathfrak{A} = \sigma(E(X \mid \mathfrak{A}))$. Let us put $\mathscr{B} = \sigma(\sigma(X) \cup \mathfrak{A})$. To use Lemma 4, we take (finite) σ-fields $\mathfrak{A} = \mathfrak{A}_0 \subset \ldots \subset \mathfrak{A}_n = \mathscr{B}$ in such a way that for fixed i, $1 \leqslant i \leqslant n$, there exists a partition $A_1, \ldots, A_m$ of Ω satisfying

$$\mathfrak{A}_i = \sigma(A_1, \ldots, A_m), \quad \mathfrak{A}_{i-1} = \sigma(A_1 \cup A_2, A_3, \ldots, A_m).$$

Then $F_{E(X \mid \mathfrak{A}_{i-1})} \in \mathscr{S}(F_{E(X \mid \mathfrak{A}_i)})$ by Lemma 4. Thus, $F_{E(X \mid \mathfrak{A})}$ belongs to $\mathscr{S}(F_X)$ by Remark 3.

LEMMA 6. *According to* (1) *and* (2), *the relations $G \in \mathscr{C}(F)$ and $H \in \mathscr{C}(G)$ imply $H \in \mathscr{C}(F)$.*

Proof. By assumption, $F = F_X$, $G = F_{E(X \mid Y)}$, $G = F_{\tilde{X}}$, $H = F_{E(\tilde{X} \mid \tilde{Y})}$, and one can assume that random variables X, Y, $\tilde{X}$, $\tilde{Y}$ are defined on the same probability space and that $E(X \mid Y) = Y$ and $E(\tilde{X} \mid \tilde{Y}) = \tilde{Y}$. On $\boldsymbol{R}^3$, there exists a distribution $d^{(3)}$ satisfying

$$d^{(3)}_{X^1} = p_{\tilde{X}}, \quad d^{(3)}(X^2 \mid X^1 = x^1) = P(\tilde{X} \mid \tilde{Y} = x^1),$$

$$d^{(3)}(X^3 \mid X^1 = x^1,\, X^2 = x^2) = P(X \mid Z = x^2).$$

Then $d^{(3)}_{X^3} = X$, $E_{d^{(3)}}(X^3 \mid X^1) = X^1$, $d^{(3)}_{X^1} = p_{\tilde{Y}}$ and the proof is complete.

LEMMA 7. *Assume that a random variable X is defined on a probability space without atoms and that $X = a$ on A and $X = d$ on D for some numbers $a < b \leq c < d$ and events A and D. Then, for some partition B, C of the event $A \cup D$ with $P(C) = P(A)$ and $P(B) = P(D)$, we have $E(X \mid B) = b$ or $E(X \mid C) = c$.*

The proof goes by an elementary calculation.

LEMMA 8. *According to notation* (1) *and* (2), *$\mathscr{S}(F)$ is contained in $\mathscr{C}(F)$ for $F \in \mathscr{S}$.*

Proof. Let $G \in \mathscr{S}(F)$ for some distribution functions F, $G \in \mathscr{S}$. Let us put

$$a = \sup\{x;\ F(t) = G(t) \text{ for } t < x\}, \quad d = \inf\{x;\ F(t) = G(t) \text{ for } t > x\},$$

$$b = \sup\{x;\ F, G \text{ are constant on } (a, b)\},$$

$$c = \inf\{x;\ F, G \text{ are constant on } (c, d)\}.$$

Let X be a random variable defined on a probability space without atoms, satisfying $F_X = F$. Obviously, there exist events $A \subset (X = a)$ and $D \subset (X = d)$ satisfying

$$P(A) = F(t) - G(t) \quad \text{for } t \in (a, b),$$

$$P(D) = G(t) - F(t) \quad \text{for } t \in (c, d).$$

Let a σ-field $\mathfrak{A}$ be generated by events $(X = x) \cap (A \cup D)^c$ for $x \in \mathbf{R}$, and events B and C be defined as in Lemma 7. Then the distribution function $F_1 = F_{E(X \mid \mathfrak{A})}$ satisfies

$$F_1 \in \mathscr{C}(F), \quad G \in \mathscr{S}(F_1),$$

$$G(t) = F_1(t) \quad \text{for } t \in (-\infty, a_1) \cup (d_1, \infty),$$

$$F(t) = F_1(t) \quad \text{for } t \in (a_1, d_1),\ a_1 \geqslant a,\ d_1 \leqslant d,$$

and $a_1 = b$ or $d_1 = c$.

In consequence, one can obtain a sequence of distribution functions $F_0, \ldots, F_n$ satisfying $F_0 = F$, $F_n = F$, $F_i \in \mathscr{C} F_{i-1}$, $i = 1, \ldots, n$. By Lemma 6, the proof is completed.

Our Theorem 1 is a consequence of Theorem 1 and Lemmas 5 and 8.

REFERENCES

[1] J. Jakubowski and S. Kwapień, *On multiplicative systems of functions*, Bull. Acad. Polon. Sci. 27 (1979), pp. 689–694.

[2] J. Karamata, *Sur une inégalité relative aux fonctions convexes*, Publ. Math. Univ. Belgrade 1 (1932), pp. 145–148.

[3] W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York 1979.

[4] P. A. Meyer, *Probability and Potentials*, Blaisdell Publishing Company, 1966.

[5] H. V. Weizsacker and G. Winkler, *Non-compact extremal integral representations: some probabilistic aspects*, in: *Functional Analysis: Surveys and Recent Results 2*, K.-D. Bierstedt and B. Fuchssteiner (Eds.), North-Holland Publishing Company, 1980.

Faculty of Mathematics
University of Łódź
Banacha 22
PL-90-238 Łódź, Poland
E-mail: adampasz@math.uni.lodz.pl

Institute of Mathematics
Polish Academy of Sciences
Śniadeckich 8
PL-00-950 Warszawa, P.O. Box 137, Poland

Received on 8.5.2000