The Annals of Probability
1979, Vol. 7, No. 6, 1046–1050

CONDITIONAL DISTRIBUTIONS AS DERIVATIVES

BY J. PFANZAGL

University of Cologne

Let $(X, \mathscr{A}, P)$ be a probability space, Y a complete separable metric space, Z a separable metric space, and $s : X \to Y$, $t : X \to Z$ Borel measurable functions. Then the weak limit of $P\{s \in B, t \in C\}/P\{t \in C\}$ for $C \downarrow \{z\}$ exists for P-a.a. $z \in Z$, and is a regular conditional distribution of s, given t.

Let $(X, \mathscr{A})$, $(Y, \mathscr{B})$, $(Z, \mathscr{C})$ be measurable spaces, $s : X \to Y$ and $t : X \to Z$ measurable maps, and $P : \mathscr{A} \to [0, 1]$ a probability measure. We shall try to put into mathematical terms the intuitive notion of the conditional distribution of s, given $t(x) = z$. The most natural definition would be: the conditional distribution assigns

$$B \to \frac{P\{x \in X : s(x) \in B, t(x) = z\}}{P\{x \in X : t(x) = z\}}.$$

Since $P(t^{-1}\{z\}) > 0$ is not necessarily true in general, this definition is meaningless and the conditional distribution has to be defined in a different way, for instance as the limit of

$$\frac{P\{x \in X : s(x) \in B, t(x) \in C\}}{P\{x \in X : t(x) \in C\}} \quad \text{as} \quad C \downarrow \{z\}.$$

This, however, requires topological assumptions on $(Z, \mathscr{C})$ which make the concept of $C\downarrow\{z\}$ meaningful, and which guarantee that a density can be obtained as a derivative. For this reason the usual approach to conditional distributions is different. Instead of defining the conditional distribution given $t(x) = z$ as the primary object one considers the conditional expectation of $1_{s^{-1}B}$, given t, say φ_B, for each $B \in \mathscr{B}$ separately, where $\varphi_B : Z \to [0, 1]$ is defined by

$$\int \varphi_B(z) 1_C(z) P{*}t(dz) = P(s^{-1}B \cap t^{-1}C) \qquad \text{for all} \quad C \in \mathscr{C}.$$

(The induced probability measure $P{*}t|\mathscr{C}$ is defined by $P{*}t(C) := P(t^{-1}C)$, $C \in \mathscr{C}$.) Considering $\varphi_B(z)$ for fixed z as a function of B one finds that it will be a content for $P{*}t$ a.a. $z \in Z$ if $\mathscr{B}$ is countably generated. Under suitable regularity conditions (e.g., if $(Y, \mathscr{B})$ is a complete separable metric space, endowed with its Borel field), the functions φ_B, being unique up to $P{*}t$-equivalence only, can be chosen such that $B \to \varphi_B(z)$ is a probability measure for $P{*}t$-a.a. $z \in Z$. An example of Dieudonné (see Doob (1953), page 624) shows that this is not possible in general.

Received March 24, 1978.

AMS 1970 *subject classifications.* Primary 60A10; secondary 28A15.

Key words and phrases. Conditional distributions, differentiation of measures.

The shortcoming of this approach: since the conditional distributions thus defined are unique up to $P*t$-null sets only, it is, in general, not meaningful to speak of the conditional distribution of s, given a fixed value $t(x) = z$. Unless $P\{x : t(x) = z\} > 0$ one can always change the definition of φ_B at z without changing its character as a conditional distribution. For this reason, the local approach, defining conditional distributions as limits, has recently been emphasized by Tjur (1974). His approach is, however, based on Radon measures, and his emphasis is on conditional distributions having nice properties. (His definition of a conditional distribution given $t(x) = z$ (see page 20), for instance, guarantees its continuity in z.) The price: there is no reasonably general existence theorem for such conditional distributions.

In this note we shall present a local approach to conditional distributions based on differentiation of measures and the classical concept of a conditional distribution. Our theorem generalizes a result of Kolmogoroff (1933, page 45) from real valued functions t, s to more general ones.

In the following we need the concept of a filterbase. Recall that a filterbase $\mathcal{F}$ is a system of nonvoid subsets of a set Z with the property that the intersection of any two sets of $\mathcal{F}$ contains a set of $\mathcal{F}$. The filterbase $\mathcal{F}$ converges to a point $z \in Z$ if every neighborhood of z contains a set of $\mathcal{F}$.

Given a system $\mathcal{H}$ of sets let $\mathcal{H}_z$ denote the class of all sets in $\mathcal{H}$ containing $z \in Z$.

DEFINITION. A D-space is a Hausdorff space Z endowed with a system $\mathcal{H}$ of Borel sets such that

(i) For each $z \in Z$, $\mathcal{H}_z$ is a filterbase converging to z.

(ii) Let P and Q be two probability measures defined on $\mathcal{C}^*$, the completion of the Borel σ-field with respect to P, such that $Q|\mathcal{C}^* \ll P|\mathcal{C}^*$. Then the net $(Q(C)/P(C))_{C \in \mathcal{H}_z}$ (for $P(C) = 0$ define $Q(C)/P(C) = 0$) converges for P-a.a. $z \in Z$. Considered as a function of z, this limit is a density of $Q|\mathcal{C}^*$ with respect to $P|\mathcal{C}^*$.

CONVENTION. If a function $F|\mathcal{H}$ converges for the filterbase $\mathcal{H}_z \subset \mathcal{H}$ then this limit will be denoted by $\lim_{C \to z} F(C)$.

It is well known (see, e.g., Hahn-Rosenthal (1948), page 254, 17.3.1, 17.2.63, and page 253, 17.2.61) that for any separable metric space Z there is a Vitali system $\mathcal{H}$ of Borel sets such that $(Z, \mathcal{H})$ is a D-space. In the particular case of $Z = \mathbb{R}^k$ one can take for $\mathcal{H}$ the class of all open cubes.

For each $C \in \mathcal{C}$ with $P(t^{-1}C) > 0$ let

$$(1) \qquad M_C(B) := \frac{P(s^{-1}B \cap t^{-1}C)}{P(t^{-1}C)}, \qquad B \in \mathcal{B}.$$

We remark that $M_C|\mathcal{B}$ is a probability measure.

If the net of probability measures $(M_C|\mathcal{B})_{C \in \mathcal{H}_z}$ converges weakly, the limit, being a probability measure on $\mathcal{B}$, will be denoted by $M(z, \cdot)|\mathcal{B}$. (With a stronger type of convergence, e.g., setwise convergence of $M_C|\mathfrak{B}$ to $M(z, \cdot)|\mathfrak{B}$, this definition becomes too restrictive to be useful.)

The following theorem gives conditions under which $M(z, \cdot)|\mathfrak{B}$ is a conditional distribution of s, given $t(x) = z$. Notice that the conditional distribution depends on the system $\mathfrak{H}$. (To obtain a trivial example, assume that the induced probability measure $P*(s, t)$ has a density with a jump along $t(x) = z$. If we take in (1) the limit for $\Delta \to 0$ with $C = (z - a\Delta, z + b\Delta)$, then $M(z, \cdot)$ will depend on a and b. For a less trivial example see Kac and Slepian (1959), Section 2.) Depending on the system $\mathfrak{H}$, the conditional distribution is unique a.e. only. In spite of this, it adds to the intuitive interpretation of the concept of a conditional distribution that it can be obtained as a derivative.

THEOREM. *Assume that* $(Z, \mathfrak{H})$ *is a D-space and* $(Y, \mathfrak{B})$ *is a metric space endowed with its Borel field.*

(i) *If* $M(z, \cdot)|\mathfrak{B}$ *exists for* $P*t$ – a.a. $z \in Z$, *then it is a conditional probability distribution of* s, *given* $t(x) = z$, i.e., *for every* $B \in \mathfrak{B}$, $z \to M(z, B)$ *is measurable with respect to the completion of* $\mathcal{C}$ *under* $P*t$, *and*

$$\int M(z, B)1_C(z)P*t(dz) = P(s^{-1}B \cap t^{-1}C) \qquad \textit{for all} \quad C \in \mathcal{C}. \tag{2}$$

(ii) *If* $(Y, \mathfrak{B})$ *is a complete separable metric space then* $M(z, \cdot)|\mathfrak{B}$ *exists for* $P*t$ – a.a. $z \in Z$.

PROOF.

(i) Assume that $M(z, \cdot)|\mathfrak{B}$ exists for all $z \notin N$, where $P*t(N) = 0$. Let $f : Y \to \mathbb{R}$ be a bounded continuous function, and f_0 a conditional expectation of $f \circ s$ given $t(x) = z$, i.e.

$$\int f_0(\xi)1_C(\xi)P*t(d\xi) = \int f(s(x))1_C(t(x))P(dx) \quad \text{for all} \quad C \in \mathcal{C}. \tag{3}$$

By definition of M_C (see (1))

$$\int f(y)M_C(dy) = \int f(s(x))1_C(t(x))P(dx)/P(t^{-1}C).$$

Together with (3)

$$\int f(y)M_C(dy) = \int f_0(\xi)1_C(\xi)P*t(d\xi)/P*t(C). \tag{4}$$

Hence there exists a $P*t$-null set N' such that $z \notin N'$ implies

$$\begin{aligned} M(z,f) := \int f(y)M(z, dy) &= \lim_{C\to z}\int f(y)M_C(dy) \\ &= \lim_{C\to z}\int f_0(\xi)1_C(\xi)P*t(d\xi)/P*t(C) = f_0(z). \end{aligned} \tag{5}$$

(The last equality follows from the fact that $(Z, \mathfrak{H})$ is a D-space.)

Hence $z \to M(z, f)$ is measurable with respect to the completion of $\mathcal{C}$ under $P*t$, say $\mathcal{C}^*$. It is easy to see that the class of all functions f for which $z \to M(z, f)$ is $\mathcal{C}^*$-measurable, is closed under linear combinations and monotone convergence. If such a class of functions contains all bounded and continuous functions, it contains all Borel-measurable functions, if the basic space is metric. Hence $z \to M(z, B)$ is $\mathcal{C}^*$-measurable for every $B \in \mathfrak{B}$.

Now let $M'_C|\mathfrak{B}$ be defined by

$$M'_C(B) := P(s^{-1}B \cap t^{-1}C), \qquad B \in \mathfrak{B}$$

and $M''_C|\mathfrak{B}$ by

$$M''_C(B) := \int M(\xi, B)1_C(\xi)P*t(d\xi), \qquad B \in \mathfrak{B}.$$

From (3) and (5),

$$M''_C(f) = M'_C(f).$$

Since this holds true for every bounded continuous function f, it implies (see Parthasarathy (1967), page 39, Theorem 5.9) that $M''_C(B) = M'_C(B)$ for all $B \in \mathfrak{B}$, which is (2).

(ii) Since Y is a separable metric space, there exists (see Parthasarathy (1967), page 47, Theorem 6.6) a countable class of bounded and continuous functions $f_k : Y \to \mathbb{R}$ such that a net of probability measures $Q_\alpha|\mathfrak{B}$, $\alpha \in A$, converges weakly to a probability measure $Q|\mathfrak{B}$ iff $(Q_\alpha(f_k))_{\alpha \in A} \to Q(f_k)$ for every $k \in \mathbb{N}$.

Since Y is a complete separable metric space, there exists (see, e.g., Ash (1972), page 265, Theorems 6.6.5 and 6.6.6) a regular conditional distribution of s, given $t(x) = z$, say $\overline{M}|Z \times \mathfrak{B}$; i.e., $\overline{M}(z, \cdot)|\mathfrak{B}$ is a p-measure for $P*t$ – a.a. $z \in Z$ and fulfills (2).

We shall show that for $P*t$ – a.a. $z \in Z$, $(M_C|\mathfrak{B})_{C \in \mathcal{K}_z}$ converges weakly to $\overline{M}(z, \cdot)|\mathfrak{B}$.

It follows easily from (2) that

$$\int \overline{M}(\xi, f_k)1_C(\xi)P*t(d\xi) = \int f_k(s(x))1_C(t(x))P(dx)$$
$$\text{for} \quad k \in \mathbb{N} \text{ and } C \in \mathcal{C}.$$

Since $(Z, \mathcal{K})$ is a D-space,

$$\lim_{C \to z} \int \overline{M}(\xi, f_k)1_C(\xi)P*t(d\xi)/P*t(C) = \overline{M}(z, f_k)$$

for $z \notin N_k$ with $P*t(N_k) = 0$.

Hence $z \notin \cup_1^\infty N_k$ implies

$$\lim_{C \to z} M_C(f_k) = \overline{M}(z, f_k) \qquad \text{for} \quad k \in \mathbb{N}.$$

By the choice of the class $\{f_k : k \in \mathbb{N}\}$ this implies that for $z \notin \cup_1^\infty N_k$, $(M_C|\mathfrak{B})_{C \in \mathcal{K}_z}$ converges weakly to $\overline{M}(z, \cdot)|\mathfrak{B}$.

Acknowledgment. The author wishes to thank Mr. Tue Tjur, Mr. B. B. Winter, and the referee for some critical remarks.

REFERENCES

Ash, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York.

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

Hahn, H. and Rosenthal, A. (1948). *Set Functions*. Univ. New Mexico Press.

Kac, M. and Slepian, D. (1959). Large excursions of Gaussian processes. *Ann. Math. Statist.* **30** 1215–1228.

KOLMOGOROFF, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York—London.

TJUR, T. (1974). *Conditional Probability Distributions*. Lecture Notes 2, Inst. Math. Statist., Univ. Copenhagen.

MATHEMATISCHES INSTITUT
DER UNIVERSITÄT ZU KÖLN
WEYERTAL 86–90
5000 KÖLN 41
WEST GERMANY