Pavla Kunderová

On the existence of conditional density functions

Sborník prací Přírodovědecké fakulty University Palackého v Olomouci. Matematika, Vol. 17 (1978), No. 1, 115--121

1978 — ACTA UNIVERSITATIS PALACKIANAE OLOMUCENSIS
FACULTAS RERUM NATURALIUM — TOM 57

Katedra matematické analýzy a numerické matematiky přírodovědecké fakulty University Palackého v Olomouci

Vedoucí katedry: prof. RNDr. Miroslav Laitoch, CSc.

ON THE EXISTENCE OF CONDITIONAL DENSITY FUNCTIONS

PAVLA KUNDEROVÁ

(*Received March 31, 1977*)

1. Introduction

The conditional probability is generally defined thus: Let $(X, \mathscr{A}, P)$ be a probability space, $(Y, \mathscr{B})$ a measurable space and $T: X \to Y$ an arbitrary measurable mapping. Let us put

$$\nu_A(B) = P(A \cap T^{-1}(B)), \qquad A \in \mathscr{A}, B \in \mathscr{B}.$$

Then ν_A and PT^{-1} are measures on $\mathscr{B}$ where ν_A is absolutely continuous with respect to PT^{-1} (denoted by $\nu_A \ll PT^{-1}$). According to the Radon-Nikodym theorem there exists an integrable function $P(A/y)$ on Y such that

$$\nu_A(B) = P(A \cap T^{-1}(B)) = \int_B P(A/y) \, \mathrm{d}PT^{-1}(y) \tag{1}$$

holds for all sets $B \in \mathscr{B}$. The function $P(A/y)$ is uniquely determined by (1) almost everywhere with respect to PT^{-1}.

The function $P(A/y)$ is called the conditional probability of the event A under the condition of $T(x) = y$ [or the conditional probability of the event A with a given value $T(x)$]. Cf. [3], page 203.

Let E_n be an n-dimensional Euclidean space and $\mathscr{B}_n$ a σ-algebra of the Borel subsets of E_n. In case of $X = E_n$, $\mathscr{A} = \mathscr{B}_n$, $Y = E_1$, $\mathscr{B} = \mathscr{B}_1$, $T: E_n \to E_1$ being a measurable function and the probability P having the density function $f(x)$ with respect to the Lebesgue measure (if the density function with respect to the Lebesgue measure is involved in the sequel, we will not write it), the function $P(A/y)$ is uniquely determined almost everywhere with respect to PT^{-1} for the so-called "coordinate-functions" $T(x) = T(x_1, \dots, x_n) = x_i$, $(i = 1, 2, \dots, n)$ by the relation

$$P(A/y) = \int_A h(x/y) \, \mathrm{d}x, \qquad A \in \mathscr{B}_n \tag{2}$$

where

$$h(x/y) = \begin{cases} \dfrac{f(x)}{g(y)}, & \text{if } x \in T^{-1} \text{ and } g(y) > 0 \text{ as well} \\ 0, & \text{if } x \notin T^{-1}(y) \text{ or } g(y) = 0 \end{cases}$$

is the so-called conditional density function and $g(y)$ the density function of the probability distribution of $T(x)$. See [2], page 71.

Below we will deduce sufficient conditions for the existence of conditional density functions under some weakened assumptions laid on the function $T(x)$. To the proof we will use results from the theory of integrals with respect to the Hausdorff measure

2. The basic concepts and theorems

Definition 1. Let X be a separable metric space with a metric ϱ, p a natural number and $\alpha(p)$ a volume (p-dimensional Lebesgue measure) of the p-dimensional sphere with diameter one. Under the p-dimensional Hausdorff (outer) measure of the set $E \subset X$ we understand

$$H^p(E) = \sup_{\varepsilon > 0} \{\alpha(p) \inf [\sum_{i=1}^{\infty} (\text{diam } E_i)^p : E \subset \bigcup_{i=1}^{\infty} E_i, \text{diam } E_i < \varepsilon, i = 1, 2, \ldots]\}$$

with diam $E = \sup_{x,y \in E} \varrho(x, y)$, and the infimum being taken over all the coverings of the set E having the properties mentioned. (Cf. [3], page 58).

It holds

1. Every Borel set in X is H^p-measured (see [3], page 58).

2. H^n defined in E_n is identical with the Lebesgue measure in E_n (see [4], theorem 1.17).

3. If H^n is defined in E_n and $k < n$ (k, n natural), then $H^n(E_k) = 0$ (see [4], theorem 1.18).

The definition of the Hausdorff measure H^p can be also extended to $p = 0$ as follows:
the 0-dimensional Hausdorff measure $H^0(A)$ equals the number (possibly ∞) of elements of A.

In what follows we use the following notations:

H_n^m the m-dimensional Hausdorff measure defined in E_n

det $\mathbf{M}$ the determinant of the square matrix $\mathbf{M}$

$\mathbf{M}'$ the matrix transposed to $\mathbf{M}$

$f: E_n \to E_k$, $f = (f_1, \ldots, f_k)$

$\mathbf{D}f(x)$ the matrix with the elements $\dfrac{\partial f_i(x)}{\partial x_j}$, $x = (x_1, \ldots, x_n)$

$Jf(x) = \sqrt{\det \mathbf{D}f(x) (\mathbf{D}f(x))'}$.

The function $Jf(x)$ is continuous on E_n for the functions of class C_1 (i.e. when $f_j(x)$ have continuous derivatives of the first order in E_n). If $k = 1$, then $Jf(x) = |\operatorname{grad} f(x)|$. In paper [1], page 426, the following two theorems are formulated more generally.

Lemma 1. *Let $f: E_n \to E_k$ $(k \leqq n)$ be a Lipschitzian mapping of class C_1. Then for an arbitrary H_n^n-measurable set $A \subset E_n$*

$$\int_A Jf(x)\,\mathrm{d}H_n^n(x) = \int_{E_k} H_n^{n-k}(A \cap f^{-1}(y))\,\mathrm{d}H_k^k(y). \tag{3}$$

Lemma 2. *Let $f: E_n \to E_k$ $(k \leqq n)$ be a Lipschitzian mapping of class C_1 and let $g: E_n \to E_1$ be H_n^n-integrable. Then*

$$\int_{E_n} g(x)\,Jf(x)\,\mathrm{d}H_n^n(x) = \int_{E_k}\Big(\int_{f^{-1}(y)} g(x)\,\mathrm{d}H_n^{n-k}(x)\Big)\,\mathrm{d}H_k^k(y). \tag{4}$$

Remark 1. According to property 2 of the measure H_n^n the above statements (3) and (4) can be written in the following form

$$\int_A Jf(x)\,\mathrm{d}x = \int_{E_k} H_n^{n-k}(A \cap f^{-1}(y))\,\mathrm{d}y, \tag{3'}$$

$$\int_{E_n} g(x)\,Jf(x)\,\mathrm{d}x = \int_{E_k}\Big(\int_{f^{-1}(y)} g(x)\,\mathrm{d}H_n^{n-k}(x)\Big)\,\mathrm{d}y. \tag{4'}$$

Lemma 3. *Let $T: E_n \to E_1$ be a Lipschitzian function of class C_1 such that for all $x \in E_n$ we have $|\operatorname{grad} T(x)| > 0$. Next $h: E_n \to E_1$* be a *H_n^n-integrable function and let λ be a measure on $\mathscr{B}_n$ defined by the relation*

$$\lambda(B) = \int_B \frac{1}{|\operatorname{grad} T(x)|}\,\mathrm{d}H_n^{n-1}(x), \qquad B \in \mathscr{B}_n.$$

Then

$$\int_{E_n} h(x)\,\mathrm{d}x = \int_{E_1}\Big(\int_{T^{-1}(y)} h(x)\,\mathrm{d}\lambda(x)\Big)\,\mathrm{d}y. \tag{5}$$

Proof: Let us define a function $g(x)$ by the relation $h(x) = g(x)\,|\operatorname{grad} T(x)|$. By appealing to (4′) we get

$$\int_{E_n} h(x)\,\mathrm{d}x = \int_{E_n} g(x)\,|\operatorname{grad} T(x)|\,\mathrm{d}x = \int_{E_1}\Big(\int_{T^{-1}(y)} g(x)\,\mathrm{d}H_n^{n-1}(x)\Big)\,\mathrm{d}y$$

λ is completely σ-finite measure on $\mathscr{B}_n$ and $H_n^{n-1} \ll \lambda$. According to [3], theorem 2, page 133, it is possible to introduce the new measure λ in the integral on the right side, which leads to

$$\int_{E_n} h(x)\,\mathrm{d}x = \int_{E_1}\Big(\int_{T^{-1}(y)} g(x)\,|\operatorname{grad} T(x)|\,\mathrm{d}\lambda(x)\Big)\,\mathrm{d}y.$$

The statement is an immediate consequence.

3. Theorems on the existence of conditional density functions

Theorem 1. *Let $(E_n, \mathscr{B}_n, P)$ be a probability space, $(E_1, \mathscr{B}_1)$ a measurable space and $T: E_n \to E_1$ a Lipschitzian function of class C_1 such that $|\operatorname{grad} T(x)| > 0$ for all $x \in E_n$. Next let $f(x)$ be a density function of the probability P and*

$$g(y) = \int_{T^{-1}(y)} f(x)\, \mathrm{d}\lambda(x),$$

with λ being defined in the same way as in Lemma 3.
Then for the conditional probability $P(A/y)$

$$P(A/y) = \begin{cases} \dfrac{1}{g(y)} \displaystyle\int_{A \cap T^{-1}(y)} \dfrac{f(x)}{|\operatorname{grad} T(x)|} \mathrm{d}H_n^{n-1}(x), & \text{if } g(y) > 0 \\ 0, & \text{if } g(y) = 0. \end{cases} \tag{6}$$

Proof: According to (5) we have for $A \in \mathscr{B}_n$

$$P(A) = \int_A f(x)\, \mathrm{d}x = \int_{E_n} \chi_A(x) f(x)\, \mathrm{d}x = \int_{E_1} \Big(\int_{A \cap T^{-1}(y)} f(x)\, \mathrm{d}\lambda(x) \Big) \mathrm{d}y \tag{7}$$

and thus for $A \in \mathscr{B}_n$, $B \in \mathscr{B}_1$

$$P(A \cap T^{-1}(B)) = \int_{A \cap T^{-1}(B)} f(x)\, \mathrm{d}x = \int_B \Big(\int_{A \cap T^{-1}(y)} f(x)\, \mathrm{d}\lambda(x) \Big) \mathrm{d}y =$$

$$= \int_B \left(\int_{A \cap T^{-1}(y)} f(x) \frac{1}{|\operatorname{grad} T(x)|} \mathrm{d}H_n^{n-1}(x) \right) \mathrm{d}y \tag{8}$$

because (analogous to the proof of Lemma 3) the conditions for introducing a new measure in the last integral are satisfied. Due to (7) we have for an arbitrary $B \in \mathscr{B}_1$

$$PT^{-1}(B) = \int_{T^{-1}(B)} f(x)\, \mathrm{d}x = \int_B \Big(\int_{T^{-1}(y)} f(x)\, \mathrm{d}\lambda(x) \Big) \mathrm{d}y = \int_B g(y)\, \mathrm{d}y.$$

a) Suppose that $g(y) > 0$. Including the new measure PT^{-1} in the last integral (conditions satisfied), we obtain

$$P(A \cap T^{-1}(B)) = \int_B \left(\int_{A \cap T^{-1}(y)} \frac{f(x)}{|\operatorname{grad} T(x)|} \mathrm{d}H_n^{n-1}(x) \right) \frac{1}{g(y)} \mathrm{d}PT^{-1}(y)$$

wherefrom by definition of the conditional probability the statement (6) follows.

b) If $g(y) = \int_{T^{-1}(y)} f(x)\, \mathrm{d}\lambda(x) = 0$, the statement is evident in respecting $f(x) \geqq 0$ and

$$P(A \cap T^{-1}(B)) = \int_B \Big(\int_{A \cap T^{-1}(y)} f(x)\, \mathrm{d}\lambda(x) \Big) \mathrm{d}y.$$

Corollary 1. *If the assumptions of Theorem 1 are satisfied, the conditional probability $P(A/y)$ of the event A under the condition of $T(x) = y$ has the conditional density function with respect to the measure H_n^{n-1} stated below*

$$p(x/y) = \begin{cases} \dfrac{f(x)}{g(y)\,|\,\mathrm{grad}\,T(x)\,|}, & \textit{if } x \in T^{-1}(y) \textit{ and } g(y) > 0 \textit{ as well} \\ 0, & \textit{if } x \notin T^{-1}(y) \textit{ or } g(y) = 0 \end{cases}$$

which immediately follows from (6).

Remark 2. The functions of the type $T(x) = T(x_1, \ldots, x_n) = x_i$, $(i = 1, 2, \ldots, n)$, mentioned in the introduction, satisfy the conditions of Theorem 1 and the expression of $P(A/y)$ in (2) represents a special case of (6).

Theorem 2. *Let* $(E_n, \mathscr{B}_n, P)$ *be a probability space, and* $(E_k, \mathscr{B}_k)$ *be a measurable space* $(k \leqq n)$. *Next let* $T: E_n \to E_k$ *be a Lipschitzian mapping of class* C_1 *such that* $JT(x) \neq 0$ *everywhere in* E_n. *Let P have a density function* $f(x)$ *and let*

$$g(y) = \int_{T^{-1}(y)} f(x)\,\mathrm{d}\lambda(x)$$

where

$$\lambda(C) = \int_C \frac{1}{JT(x)}\,\mathrm{d}H_n^{n-k}(x), \qquad \textit{for every } C \in \mathscr{B}_n.$$

Then

$$P(A/y) = \begin{cases} \dfrac{1}{g(y)} \displaystyle\int_{A \cap T^{-1}(y)} \frac{f(x)}{JT(x)}\,\mathrm{d}H_n^{n-k}(x), & \textit{if } g(y) > 0 \\ 0, & \textit{if } g(x) = 0 \end{cases} \tag{9}$$

Proof: In analogy with the proof of Lemma 3 we will show that foı any H_n^n-integrable function $h: E_n \to E_1$ we can write

$$\int_{E_n} h(x)\,\mathrm{d}x = \int_{E_k} \Big(\int_{T^{-1}(y)} h(x)\,\mathrm{d}\lambda(x) \Big)\,\mathrm{d}y.$$

Thus for any $A \in \mathscr{B}_n$

$$P(A) = \int_A f(x)\,\mathrm{d}x = \int_{E_n} \chi_A(x) f(x)\,\mathrm{d}x = \int_{E_k} \Big(\int_{A \cap T^{-1}(y)} f(x)\,\mathrm{d}\lambda(x) \Big)\,\mathrm{d}y \tag{10}$$

and consequently for $A \in \mathscr{B}_n$, $B \in \mathscr{B}_k$ we obtain

$$\begin{aligned} P(A \cap T^{-1}(B)) &= \int_{A \cap T^{-1}(B)} f(x)\,\mathrm{d}x = \int_{E_k} \Big(\int_{A \cap T^{-1}(B) \cap T^{-1}(y)} f(x)\,\mathrm{d}\lambda(x) \Big)\,\mathrm{d}y = \\ &= \int_B \Big(\int_{A \cap T^{-1}(y)} f(x)\,\mathrm{d}\lambda(x) \Big)\,\mathrm{d}y = \int_B \left(\int_{A \cap T^{-1}(y)} \frac{f(x)}{JT(x)}\,\mathrm{d}H_n^{n-k}(x) \right) \mathrm{d}y. \end{aligned} \tag{11}$$

The introducing of measure H_n^{n-k} in the last integral is justified since λ is completely σ-finite measure and $H_n^{n-k} \ll \lambda$. Following (10) we can write for any $B \in \mathscr{B}_k$

$$\begin{aligned} PT^{-1}(B) &= \int_{T^{-1}(B)} f(x)\,\mathrm{d}x = \int_{E_k} \Big(\int_{T^{-1}(B) \cap T^{-1}(y)} f(x)\,\mathrm{d}\lambda(x) \Big)\,\mathrm{d}y = \\ &= \int_B \Big(\int_{T^{-1}(y)} f(x)\,\mathrm{d}\lambda(x) \Big)\,\mathrm{d}y = \int_B g(y)\,\mathrm{d}y. \end{aligned}$$

Let us assume now that $g(y) > 0$. Introducing the new measure PT^{-1} into the last integral of (11) (conditions satisfied, see [3], page 133), results in

$$P(A \cap T^{-1}(B)) = \int_B \left(\int_{A \cap T^{-1}(y)} \frac{f(x)}{JT(x)} \mathrm{d}H_n^{n-k}(x) \right) \frac{1}{g(y)} \mathrm{d}PT^{-1}(y),$$

which finally proves our statement due to definition of $P(A/y)$. If $g(y) = \int_{T^{-1}(y)} f(x) \, \mathrm{d}\lambda(x) = 0$, the statement is evident since $f(x) \geqq 0$ and we know from (11), that $P(A \cap T^{-1}(B)) = \int_B (\int_{A \cap T^{-1}(y)} f(x) \, \mathrm{d}\lambda(x)) \, \mathrm{d}y$. Similarly to Theorem 1, we obtain from Theorem 2

Corollary 2. *If the conditions of Theorem 2 are satisfied, the conditional probability $P(A/y)$ under the condition of $T(x) = y$ has the following conditional density function (with respect to the measure H_n^{n-k})*

$$p(x/y) = \begin{cases} \dfrac{f(x)}{g(y) \, JT(x)}, & \text{if } x \in T^{-1}(y) \text{ along with } g(y) > 0 \\ 0, & \text{if } x \notin T^{-1}(y) \text{ or } g(y) = 0. \end{cases}$$

Remark 3. The usually considered vector functions $T(x)$ (see i.e. [2], page 72) $T(x) = [T_1(x), \ldots, T_k(x)]$, with $T_j(x_1, \ldots, x_n) = x_{i_j}$ $(j = 1, \ldots, k;\ 1 \leqq i_j \leqq n$, i_j being an integer) are special cases of the mapping $T(x)$ of Theorem 2 and the usually presented conditional density function

$$h(x/y) = \begin{cases} \dfrac{f(x)}{g(y)}, & \text{if } x \in T^{-1}(y) \text{ and } g(y) > 0 \text{ as well} \\ 0, & \text{if } x \notin T^{-1}(y) \text{ or } g(y) = 0 \end{cases}$$

wherein $g(y)$ represents the marginal density function of the probability distribution of the mapping $T(x)$, is a special case of the density function $p(x/y)$ from Corollary 2.

REFERENCES

[1] Federer H.: *Curvature measures.* Trans. Amer. Math. Soc. 93, 426—447 (1959).

[2] Fisz M.: *Wahrscheinlichkeitsrechnung und mathematische Statistik.* Berlin 1971.

[3] Halmos P. R.: *Teorija mery.* Moskva 1953.

[4] Král J., Mařík J.: *Integrace podle Hausdorffovy míry na hladké ploše.* Časopis pro pěstování matematiky 89, 433—448 (1964).

SOUHRN

POZNÁMKA K EXISTENCI PODMÍNĚNÝCH HUSTOT

PAVLA KUNDEROVÁ

V článku se řeší problém výpočtu podmíněné hustoty pravděpodobnosti některého jevu při podmiňujícím zobrazení $T: E_n \to E_1$ (resp. $E_n \to E_k$, $k \leqq n$). Je-li T projekce do některé souřadnice, existuje klasická formule (2). Věty 1, 2 uvádějí postačující podmínky pro existenci podmíněných hustot za slabších předpokladů o zobrazení T· V důkazech jsou užity výsledky teorie křivkového integrálu podle Hausdorffovy míry

РЕЗЮМЕ

ЗАМЕЧАНИЕ К СУЩЕСТВОВАНИЮ УСЛОВНЫХ ПЛОТНОСТЕЙ

ПАВЛА КУНДЕРОВА

В работе решается проблема вычисления условной плотности вероятности некоторого события при условии $T(y) = x$, $T: E_n \to E_1$ (или $E_n \to E_k$, $k \leqq n$). Если T проекция на некоторую координату, существует классическая формула (2). В теоремах 1, 2 даны достаточные условия для существования условных плотностей в случае более слабых предположений о изображении T. В доказательствах используются утверждения теория криволинейного интеграла по мере Хаусдорффа.