11-1990

Characterization of Joint Density by Conditional Densities

A.K. Gupta
Bowling Green State University

T. Varga
Rose-Hulman Institute of Technology

MSTR 90-07

CHARACTERIZATION OF JOINT DENSITY BY CONDITIONAL DENSITIES

A.K. Gupta and T. Varga

MS TR 90-07

November 1990

Department of Mathematics
Rose-Hulman Institute of Technology
Terre Haute, IN 47803

FAX(812) 877-3198

Phone: (812) 877-8391

CHARACTERIZATION OF JOINT DENSITY BY CONDITIONAL DENSITIES

A. K. Gupta and T. Varga*

Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, Ohio 43403-0221

Key Words and Phrases: joint density; conditional density; bivariate normal density.

ABSTRACT

In this paper the relationship between joint density and conditional densities is studied. An explicit formula is given for obtaining the joint density from the conditional ones. It is illustrated for the case of bivariate normal distribution.

1. INTRODUCTION

The question of how a bivariate probability density function can be determined if the conditional density functions are given has been studied by various authors. Brucker (1979) examined the case of normal conditional densities. The conditions of his theorem were weakened by Fraser and Streit (1980). A general result about the compatibility of conditional densities was given by Abrahams and Thomas (1984). In the following paper we examine the more general case, when the conditional densities are multivariate and the dominating measure is arbitrary, not necessarily the Lebesgue measure. We give necessary and sufficient conditions for two functions to be the conditional densities of a probability density function. We also give condition under which the conditional densities determine the joint density uniquely.

*Now at Rose-Hulman Institute of Technology.

2. MAIN RESULTS

The first theorem gives a condition under which the conditional densities determine the joint density uniquely and also shows how the joint density can be obtained from the conditional densities.

THEOREM 2.1. *Let us assume that an* $n + m$ *dimensional distribution is dominated by a measure* σ *on* $\mathbb{R}^{n+m}$, *i.e.*

$$P(B) = \int_B f(z)\, d\sigma(z) \textit{ for every } B \textit{ Borel set in } \mathbb{R}^{n+m}.$$

Suppose that $\sigma = \mu \times \lambda$ *where* μ *is a measure on* $\mathbb{R}^n$ *and* λ *is a measure on* $\mathbb{R}^m$. *Assume that there exists an* $A \subset \mathbb{R}^{n+m}$ *such that* $P(A) = 1$ *and if* $z = (x,y) \in A$ *then the conditional densities* $f_1(x \mid y)$ *and* $f_2(y \mid x)$ *exist. Define*

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^m,\ (x,y) \in A\}$$
$$\mathcal{Y} = \{y \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n,\ (x,y) \in A\}.$$

(1) *Assume that there exists an* $x_0 \in \mathbb{R}^n$ *such that for every* $y \in \mathcal{Y}$, $f_1(x_0 \mid y) > 0$.

Then for $z = (x,y) \in A$ *we have*

$$(2) \qquad f(x,y) = K(x_0) \frac{f_1(x \mid y) f_2(y \mid x_0)}{f_1(x_0 \mid y)}$$

where $K(x_0) = \dfrac{1}{\displaystyle\int_A \frac{f_1(x \mid y) f_2(y \mid x_0)}{f_1(x_0 \mid y)}\, d\sigma(x,y)}$.

Proof: Let $p_1(x) = \int_{\mathbb{R}^m} f(x,y)d\lambda(y)$ and $p_2(y) = \int_{\mathbb{R}^n} f(x,y)d\mu(x)$ be the marginal densities of f. Then

$$(3) \qquad \frac{f_1(x \mid y)f_2(y \mid x_0)}{f_1(x_0 \mid y)} = \frac{\frac{f(x,y)}{p_2(y)} \cdot \frac{f(x_0,y)}{p_1(x_0)}}{\frac{f(x_0,y)}{p_2(y)}} = \frac{f(x,y)}{p_1(x_0)}.$$

Integrating (2) on A with respect to σ we obtain

$$(4) \qquad \int_A \frac{f_1(x \mid y)f_2(y \mid x_0)}{f_1(x_0 \mid y)} d\sigma(x,y) = \frac{\int_A f(x,y)d\sigma(x,y)}{p_1(x_0)} = \frac{1}{p_1(x_0)}.$$

From (3) and (4) we obtain $K(x_0)$.

REMARK 2.1. If there exists a $y_0 \in \mathbb{R}^m$ such that for every $x \in \mathcal{X}$, $f_2(y_0 \mid x) > 0$ then for $z = (x,y) \in A$

$$f(x,y) = M(y_0) \cdot \frac{f_2(y \mid x)f_1(x \mid y_0)}{f_2(y_0 \mid x)}$$

where

$$M(y_0) = \frac{1}{\int_A \frac{f_2(y \mid x)f_1(x \mid y_0)}{f_2(y_0,x)} d\sigma(x,y)}.$$

This result can be obtained in the same way as Theorem 2.1.

REMARK 2.2. It follows from Theorem 2.1 that we can obtain f(z) for every $z \in A$ if we know $f_1(x \mid y)$ and $f_2(y \mid x)$ for every $(x,y) \in A$. Moreover, it is sufficient to know $f_1(x \mid y)$ for every $(x,y) \in A$ and $f_2(y \mid x_0)$ for every $y \in \mathcal{Y}$. Similarly we can determine $f(z)$ from the knowledge of $f_2(y \mid x)$ for every $(x,y) \in A$ and $f_1(x \mid y_0)$ for every $x \in \mathcal{X}$.

REMARK 2.3. It follows from (4) that the marginal density at x_0 is

$$p_1(x_0) = \frac{1}{\displaystyle\int_A \frac{f_1(x \mid y) f_2(y \mid x_0)}{f_1(x_0 \mid y)} \, d\sigma(x,y)} .$$

REMARK 2.4. If condition (1) is not satisfied, the conditional densities may not determine the joint density uniquely as the following example shows.

Let $n = m = 1$, $\mu = \lambda =$ Lebesgue measure and $f^a(z)$ be defined as follows

$$f^a(z) = f^a(x,y) = \begin{cases} a \text{ if } (x,y) \in [0,\frac{1}{2}] \times [0,\frac{1}{2}] \\ b \text{ if } (x,y) \in (\frac{1}{2},1] \times (\frac{1}{2},1] \\ 0 \text{ elsewhere} \end{cases}$$

where $0 < a < 4$ and $a + b = 4$. Then f^a is a probability density function.

Let $A = [0,1] \times [0,1]$, hence $\mathcal{X} = [0,1]$, $\mathcal{Y} = \{0,1]$. Then

$$f_1^a(x \mid y) = f_2^a(y \mid x)$$

$$= \begin{cases} 2 \text{ if } (x,y) \in [0,\frac{1}{2}] \times [0,\frac{1}{2}] \text{ or } (x,y) \in (\frac{1}{2},1] \times (\frac{1}{2},1] \\ 0 \text{ if } (x,y) \in [0,\frac{1}{2}] \times (\frac{1}{2},1] \text{ or } (x,y) \in (\frac{1}{2},1] \times [0,\frac{1}{2}]. \end{cases}$$

Therefore the conditional probability density functions are the same for any $0 < a < 4$; hence they do not determine the joint density of x and y uniquely.

EXAMPLE 2.1. Let $f(z)$ be the bivariate normal density

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x-\mu_x}{\sigma_x}\frac{y-\mu_y}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}.$$

Then

$$f_1(x\,|\,y) = \frac{1}{\sqrt{2\pi}\,\sigma_x\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x-\mu_x - \frac{\rho\sigma_x}{\sigma_y}(y-\mu_y)\right]^2\right\}$$

$$f_2(y\,|\,x) = \frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left[y-\mu_y - \frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)\right]^2\right\}$$

and an easy computation shows that

$$\frac{f_1(x\,|\,y)f_2(y\,|\,x_0)}{f_1(x_0\,|\,y)} = \frac{1}{\frac{1}{\sqrt{2\pi}\,\sigma_x}\exp\left\{-\frac{1}{2}\left(\frac{x_0-\mu_x}{\sigma_x}\right)^2\right\}} f(x,y).$$

Therefore

$$K(x_0) = \frac{1}{\sqrt{2\pi}\,\sigma_x \int_{\mathbb{R}^2} f(x,y)dxdy} \exp\left\{-\frac{1}{2}\left(\frac{x_0-\mu_x}{\sigma_x}\right)^2\right\}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x} \exp\left\{-\frac{1}{2}\left(\frac{x_0-\mu_x}{\sigma_x}\right)^2\right\}$$

and $f(x) = K(x_0)\,\dfrac{f_1(x \mid y) f_2(y \mid x_0)}{f_1(x_0 \mid y)}$.

The second theorem gives necessary and sufficient conditions for two functions to be the conditional densities of a bivariate probability density function.

THEOREM 2.2. *Let μ be a measure on $\mathbb{R}^n$ and λ a measure on $\mathbb{R}^m$. Let $g(z)$ and $h(z)$ be nonnegative functions on $A \subset \mathbb{R}^{n+m}$, measurable with respect to the product measure $\sigma = \mu \times \lambda$. Define*

$$\mathcal{X}(y) = \{x \in \mathbb{R}^n \mid (x,y) \in A\},\ \ \mathcal{Y}(x) = \{y \in \mathbb{R}^m \mid (x,y) \in A\},$$
$$\mathcal{X} = \bigcup_{y \in \mathbb{R}^m} \mathcal{X}(y),\ \ \text{and}\ \ \mathcal{Y} = \bigcup_{x \in \mathbb{R}^n} \mathcal{Y}(x).$$

Assume that there exists an $(x_0,y_0) \in \mathbb{R}^{n+m}$ such that for every $y \in \mathcal{Y}(x_0,y) \in A$, $g(x_0,y) > 0$ and for every $x \in \mathcal{X}$, $(x,y_0) \in A$, $h(x,y_0) > 0$. Then there exists a unique probability density function f on $\mathbb{R}^{n+m}$ such that $f(z) = 0$ for $z \notin A$ and $g(z) = g(x,y)$ is the conditional probability density function $f_1(x \mid y)$ and $h(z) = h(x,y)$ is the conditional probability density function $f_2(y \mid x)$ at the points $z \in A$ if and only if the following conditions hold:

a) $\displaystyle\int_{\mathcal{X}(y)} g(x,y)\,d\mu(x) = 1$ *for* $y \in \mathcal{Y}$.

b) $\displaystyle\int_{\mathcal{Y}(x)} h(x,y)\,d\lambda(y) = 1$ *for* $x \in \mathcal{X}$.

c) $\displaystyle 0 < \int_{\mathcal{Y}} \frac{h(x_0,y)}{g(x_0,y)}\,d\lambda(y) < \infty.$

d) $$g(x,y)\cdot\frac{h(x_0,y)}{g(x_0,y)} = c(x_0,y_0)h(x,y)\frac{g(x,y_0)}{g(x,y_0)} \quad \textit{for } (x,y)\in A$$

where $c(x_0,y_0)$ *is a nonzero constant depending on* x_0 *and* y_0 *only.*

Proof; First we prove the necessity of the conditions. Let

$p_1(x) = \int_{\mathbb{R}^m} f(x,y)d\lambda(y)$ and $p_2(y) = \int_{\mathbb{R}^n} f(x,y)d\mu(x)$ be the marginal densities of f.

a) $$\int_{\mathcal{X}(y)} g(x,y)d\mu(x) = \int_{\mathcal{X}(y)} f_1(x\mid y)d\mu(x) = 1.$$

b) $$\int_{\mathcal{Y}(x)} h(x,y)d\lambda(y) = \int_{\mathcal{Y}(x)} f_2(y\mid x)d\lambda(y) = 1.$$

c) $$\int_{\mathcal{Y}} \frac{h(x_0,y)}{g(x_0,y)}d\lambda(y) = \int_{\mathcal{Y}} \frac{f_2(y\mid x_0)}{f_1(x_0\mid y)}d\lambda(y) = \int_{\mathcal{Y}} \frac{\frac{f(x_0,y)}{p_1(x_0)}}{\frac{f(x_0,y)}{p_2(y)}}d\lambda(y)$$

$$= \frac{\int_{\mathcal{Y}} p_2(y)d\lambda(y)}{p_1(x_0)} = \frac{1}{p_1(x_0)} \quad \text{and} \quad 0 < p_1(x_0) < \infty.$$

d) $$g(x,y)\frac{h(x_0,y)}{g(x_0,y)} = f_1(x\mid y)\frac{f_2(y\mid x_0)}{f_1(x_0\mid y)} = \frac{f(x,y)}{p_2(y)}\frac{\frac{f(x_0,y)}{p_1(x_0)}}{\frac{f(x_0,y)}{p_2(y)}} = \frac{f(x,y)}{p_1(x_0)},$$

$$h(x,y)\frac{g(x,y_0)}{h(x,y_0)} = f_2(y\mid x)\frac{f_1(x\mid y_0)}{f_2(y_0\mid x)} = \frac{f(x,y)}{p_1(x)}\frac{\frac{f(x,y_0)}{p_2(y_0)}}{\frac{f(x,y_0)}{p_1(x)}} = \frac{f(x,y)}{p_2(y_0)},$$

from which $c(x_0,y_0) = \frac{p_2(y_0)}{p_1(x_0)}$ follows.

Next we prove the sufficiency of the conditions. Let us define K as

$$K = \frac{1}{\int_{\mathcal{Y}} \frac{h(x_0,y)}{g(x_0,y)}d\lambda(y)}.$$

Using condition (c) K is a positive number. Let us define f as

$$f(z) = f(x,y) = \begin{cases} K \cdot g(x,y) \dfrac{h(x_0,y)}{g(x_0,y)}, & \text{for } (x,y) \in A \\ 0 & \text{elsewhere.} \end{cases}$$

Then f is nonnegative, measurable and we have to check the following properties:

i) $\displaystyle\int_{\mathbb{R}^{n+m}} f(z)d\sigma(z) = 1.$

ii) $f_1(x \mid y) = g(x,y)$ for $(x,y) \in A.$

iii) $f_2(y \mid x) = h(x,y)$ for $(x,y) \in A.$

Indeed from condition (a) we get

i)
$$\begin{aligned}\int_{\mathbb{R}^{n+m}} f(z)d\sigma(z) &= \int_A Kg(x,y)\frac{h(x_0,y)}{g(x_0,y)}d\sigma(x,y) \\ &= K \cdot \int_{\mathcal{Y}} \left(\frac{h(x_0,y)}{g(x_0,y)} \int_{\mathcal{X}(y)} g(x,y)d\mu(x)\right)d\lambda(y) \\ &= K \cdot \int_{\mathcal{Y}} \frac{h(x_0,y)}{g(x_0,y)} d\lambda(y) = 1.\end{aligned}$$

ii)
$$f_1(x \mid y) = \frac{f(x,y)}{\displaystyle\int_{\mathbb{R}^n} f(u,y)d\mu(u)} = \frac{K \cdot g(x,y)\dfrac{h(x_0,y)}{g(x_0,y)}}{\displaystyle\int_{\mathcal{X}(y)} Kg(u,y)\frac{h(x_0,y)}{g(x_0,y)}d\mu(u)}$$
$$= \frac{g(x,y)}{\displaystyle\int_{\mathcal{X}(y)} g(u,y)d\mu(u)} = g(x,y).$$

iii) Using condition (d) we have

$$f(x,y) = Kg(x,y)\frac{h(x_0,y)}{g(x_0,y)} = K \cdot c(x_0,y_0)h(x,y)\frac{g(x,y_0)}{h(x,y_0)},$$

for $(x,y) \in A$. Thus from condition (b) we get

$$f_2(y \mid x) = \frac{f(x,y)}{\int_{\mathbb{R}^m} f(x,t)d\lambda(t)} = \frac{Kc(x_0,y_0)h(x,y)\cdot\frac{g(x,y_0)}{h(x,y_0)}}{\int_{\mathcal{Y}(x)} Kc(x_0,y_0)h(x,t)\frac{g(x,y_0)}{h(x,y_0)}d\lambda(t)}$$

$$= \frac{h(x,y)}{\int_{\mathcal{Y}(x)} h(x,t)d\lambda(t)} = h(x,y).$$

The uniqueness of f follows from Theorem 2.1 since if $g(x_0,y) = f_1(x_0 \mid y) > 0$ then $f(x_0,y) > 0$ is also true.

REMARK 2.5. As a special case of Theorem 2.2, when $g \neq 0$, $h \neq 0$ on A condition (d) yields

$$\frac{g(x,y)}{h(x,y)} = c(x_0,y_0)\cdot\frac{\frac{g(x,y_0)}{h(x,y_0)}}{\frac{h(x_0,y)}{g(x_0,y)}}.$$

If we also assume that $n = m = 1$ and $\lambda = \mu =$ Lebesgue measure we obtain the result of Abrahams and Thomas (1984).

BIBLIOGRAPHY

Abrahams, J. and Thomas, J. B. (1984). A note on the characterization of bivariate densities by conditional densities. *Communications in Statistics—Theor. Meth.*, **A13**(3), 395-400.

Brucker, J. (1979). A note on the bivariate normal distribution. *Communications in Statistics—Theor. Meth.*, **A8**(2), 175-177.

Fraser, D. A. S. and Streit, F. (1980). A further note on the bivariate normal distribution. *Communications in Statistics--Theor. Meth.*, **A9**(10), 1097-1099.