The Annals of Mathematical Statistics
1968, Vol. 39, No. 4, 1236–1243

PROBABILITY DENSITIES WITH GIVEN MARGINALS[1]

BY S. KULLBACK

The George Washington University

1. Introduction. Ireland and Kullback (1968) considered the problem of estimating contingency tables with given marginals on the basis of an observed contingency table, by minimizing a discrimination information value. It was noted that the procedures they described for the case of discrete distributions may also be extended to probability densities. It is the purpose of this paper to carry out the appropriate extension. It will be noted that although the procedures and results are developed in detail for bivariate densities, as a matter of convenience, there is nothing inherent in the techniques restricting the results to bivariate densities, and indeed in Section 3 are given appropriate results for a four-variate density.

The following is the formulation of the problem to be considered. Let $\pi(x, y)$ be some bivariate probability density, and required the bivariate probability density $f(x, y)$ with given marginal probability densities $g(x)$, $h(y)$, such that

$$I(f; \pi) = \int\int f(x, y) \ln f(x, y)/\pi(x, y)\, dx\, dy \tag{1.1}$$

is a minimum over all bivariate probability densities with the given marginals. (See the discussion in Kullback [(1959), Chapter 5] and that in Good (1963), (1966) of a principle of minimum discriminability.) Note that if $I(f; \pi) < \infty$ then $f(x, y)$ determines a probability measure which is absolutely continuous with respect to the probability measure determined by $\pi(x, y)$ [Kullback, (1959), p. 5].

In order to apply the minimum discrimination information theorem [Kullback and Khairat, (1966)] to the problem formulated above define

$$T_x(\xi, \eta) = \delta(x - \xi), \qquad T_y(\xi, \eta) = \delta(y - \eta), \tag{1.2}$$

where δ is the Dirac delta-function [Rényi, (1962), p. 298] so that

$$\begin{aligned}(1.3)\quad &\int\int T_x(\xi, \eta) f(\xi, \eta)\, d\xi\, d\eta \\ &= \int\int \delta(x - \xi) f(\xi, \eta)\, d\xi\, d\eta = \int \delta(x - \xi) g(\xi)\, d\xi = g(x).\end{aligned}$$

$$\begin{aligned}(1.4)\quad &\int\int T_y(\xi, \eta) f(\xi, \eta)\, d\xi\, d\eta \\ &= \int\int \delta(y - \eta) f(\xi, \eta)\, d\xi\, d\eta = \int \delta(y - \eta) h(\eta)\, d\eta = h(y).\end{aligned}$$

By the minimum discrimination information theorem, the minimizing function is

Received July 19, 1967.

[1] Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant No. 932–65.

$$f^*(x, y) = \exp\{\int \tau(\xi)\delta(x-\xi)\,d\xi + \int \sigma(\eta)\delta(y-\eta)\,d\eta\}\pi(x, y)[M(\tau,\sigma)]^{-1} = \exp\{\tau(x) + \sigma(y)\}\pi(x, y)[M(\tau, \sigma)]^{-1} \tag{1.5}$$

where

$$M(\tau,\sigma) = \int\int \exp\{\int \tau(\xi)\delta(x-\xi)\,d\xi + \int \sigma(\eta)\delta(y-\eta)\,d\eta\}\pi(x, y)\,dx\,dy = \int\int \exp\{\tau(x) + \sigma(y)\}\pi(x, y)\,dx\,dy, \tag{1.6}$$

and τ, σ are functions to be determined so that the marginal requirements are satisfied.

By setting $\exp(\tau(x)) = a(x)$, $\exp(\sigma(y) = b(y)$ and normalizing so that $M(\tau, \sigma) = 1$, the minimizing density is

$$f^*(x, y) = a(x)b(y)\pi(x, y), \tag{1.7}$$

where $a(x)$ and $b(y)$ are functions to be determined such that

$$g(x) = a(x)\int b(y)\pi(x, y)\,dy, \qquad h(y) = b(y)\int a(x)\pi(x, y)\,dx, \qquad \int\int a(x)b(y)\pi(x, y)\,dx\,dy = 1. \tag{1.8}$$

The minimum value of (1.1) is

$$I(f^*; \pi) = \int\int f^*(x, y)\ln f^*(x, y)/\pi(x, y)\,dx\,dy = \int g(x)\ln a(x)\,dx + \int h(y)\ln b(y)\,dy. \tag{1.9}$$

A result not explicitly stated in Kullback [(1959), pp. 36–39] or [Kullback and Khairat, (1966)] but that easily follows (or that may be shown directly in this case) is

$$I(f; \pi) = I(f^*; \pi) + I(f; f^*), \tag{1.10}$$

where f is any density with the given marginals and f^* is of the form in (1.7). Since the terms in (1.10), as discrimination information numbers, are $\geqq 0$ [Kullback, (1959), pp. 14–18], it is seen that

$$I(f; \pi) \geqq I(f^*; \pi) \tag{1.11}$$

with equality, if and only if,

$$I(f; f^*) = 0; \tag{1.12}$$

that is, if and only if [Kullback, (1959), pp. 14–18],

$$f(x, y) = f^*(x, y) \quad \text{a.e..} \tag{1.13}$$

Hence, it will suffice to exhibit a density of the form of (1.7) and having the given marginals to have the minimizing density. In the next section in equations (2.18)–(2.23) is given a direct proof of the foregoing assertion as a matter of interest.

2. Iterative procedure. An iterative procedure alternately satisfying one and then the other marginal for the determination of $f^*(x, y)$ satisfying (1.7) and (1.8) will now be given and it will be shown that it converges. Iterated functions and their associated values will be indicated by subscripts. The iteration is given by

$$(2.1)\quad \begin{aligned} f_{2n-1}(x, y) &= g(x)[g_{2n-2}(x)]^{-1}f_{2n-2}(x, y), \\ f_{2n}(x, y) &= h(y)[h_{2n-1}(y)]^{-1}f_{2n-1}(x, y), \quad n = 1, 2, \cdots, f_0(x, y) = \pi(x, y). \end{aligned}$$

It is seen from (2.1) that

$$(2.2)\quad \begin{aligned} g_{2n-1}(x) &= g(x), \qquad h_{2n}(y) = h(y), \\ f_{2n-1}(x, y) &= a_n(x)b_n(y)\pi(x, y), \qquad f_{2n}(x, y) = a_n(x)b_{n+1}(y)\pi(x, y), \end{aligned}$$

so that all the iterated functions are of the form (1.7) also.

To show convergence, consider

$$(2.3)\quad \begin{aligned} I(f^*; f_{2n}) &= \int\int f^*(x, y) \ln f^*(x, y)[f_{2n}(x, y)]^{-1}\, dx\, dy \\ &= \int\int f^*(x, y) \ln f^*(x, y)[f_{2n-1}(x, y)]^{-1}\, dx\, dy \\ &\qquad - \int h(y) \ln h(y)[h_{2n-1}(y)]^{-1}\, dy \\ &= I(f^*; f_{2n-1}) - I(h; h_{2n-1}); \end{aligned}$$

and

$$(2.4)\quad \begin{aligned} I(f^*; f_{2n+1}) &= \int\int f^*(x, y) \ln f^*(x, y)[f_{2n+1}(x, y)]^{-1}\, dx\, dy \\ &= \int\int f^*(x, y) \ln f^*(x, y)[f_{2n}(x, y)]^{-1}\, dx\, dy \\ &\qquad - \int g(x) \ln g(x)[g_{2n}(x)]^{-1}\, dx \\ &= I(f^*; f_{2n}) - I(g; g_{2n}). \end{aligned}$$

Since the discrimination information values in (2.3) and (2.4) are $\geqq 0$ [Kullback, (1959), pp. 14–18], (2.3) implies

$$(2.5)\quad I(f^*; f_{2n}) \leqq I(f^*; f_{2n-1})$$

with equality, if and only if,

$$(2.6)\quad I(h; h_{2n-1}) = 0;$$

that is, if and only if [Kullback, (1959), pp. 14–18],

$$(2.7)\quad h(y) = h_{2n-1}(y) \quad \text{a.e.},$$

and (2.4) implies

$$(2.8)\quad I(f^*; f_{2n+1}) \leqq I(f^*; f_{2n})$$

with equality, if and only if,

$$(2.9)\quad I(g; g_{2n}) = 0;$$

that is, if and only if [Kullback, (1959), pp. 14–18],

$$g(x) = g_{2n}(x) \quad \text{a.e.}. \tag{2.10}$$

It is clear that (2.5) and (2.8) imply the sequence

$$I(f^*; f_1) \geqq I(f^*; f_2) \geqq \cdots \geqq I(f^*; f_{2n-1}) \geqq (f^*; f_{2n}) \geqq \cdots \geqq 0. \tag{2.11}$$

Let us first consider the case when there is equality someplace in (2.11), say

$$I(f^*; f_{2n-1}) = I(f^*; f_{2n}), \tag{2.12}$$

then from (2.5), (2.6), and (2.7), (2.12) holds, if and only if,

$$h(y) = h_{2n-1}(y) \quad \text{a.e.}. \tag{2.13}$$

From (2.1), (2.2) it then follows that

$$f_{2n}(x, y) = f_{2n-1}(x, y) \quad \text{a.e.}, \qquad g_{2n}(x) = g_{2n-1}(x) = g(x) \quad \text{a.e.}, \tag{2.14}$$

which implies

$$f_{2n+1}(x, y) = f_{2n}(x, y) \quad \text{a.e.}, \qquad h_{2n+1}(y) = h_{2n}(y) = h(y) \quad \text{a.e.} \tag{2.15}$$

and

$$I(f^*; f_{2n+1}) = I(f^*; f_{2n}), \tag{2.16}$$

and so on, that is, equality thereafter, hence

$$\begin{aligned} g_N(x) &= g(x) \quad \text{a.e.}, \qquad h_N(y) = h(y) \quad \text{a.e.}, \\ f_N(x, y) &= a_n(x) b_n(y) \pi(x, y), \qquad N \geqq 2n - 1. \end{aligned} \tag{2.17}$$

It will now be shown that $f_N(x, y) = f^*(x, y)$ a.e. (Note the last sentence of Section 1.) Since $f^*(x, y)$ minimizes (1.1) for all densities with marginals $g(x)$ and $h(y)$,

$$\begin{aligned} &\int\int f_N(x, y) \ln f_N(x, y)[\pi(x, y)]^{-1}\, dx\, dy \\ &\qquad \geqq \int\int f^*(x, y) \ln f^*(x, y)[\pi(x, y)]^{-1}\, dx\, dy \\ &\qquad = \int\int f^*(x, y) \ln f^*(x, y)[f_N(x, y)]^{-1}\, dx\, dy \\ &\qquad\quad + \int\int f^*(x, y) \ln f_N(x, y)[\pi(x, y)]^{-1}\, dx\, dy. \end{aligned} \tag{2.18}$$

But $\ln f_N(x, y)[\pi(x, y)]^{-1} = \ln a_n(x) + \ln b_n(y)$, so that

$$\begin{aligned} &\int\int f_N(x, y) \ln f_N(x, y)[\pi(x, y)]^{-1}\, dx\, dy \\ &\qquad = \int g(x) \ln a_n(x)\, dx + \int h(y) \ln b_n(y)\, dy \\ &\qquad = \int\int f^*(x, y) \ln f_N(x, y)[\pi(x, y)]^{-1}\, dx\, dy, \end{aligned} \tag{2.19}$$

and (2.18) then implies

$$0 \geqq \int\int f^*(x, y) \ln f^*(x, y)[f_N(x, y)]^{-1}\, dx\, dy = I(f^*; f_N). \tag{2.20}$$

Since $I(f^*; f_N)$ is a discrimination information value [Kullback, (1959), pp. 14–18],

$$I(f^*; f_N) \geqq 0; \tag{2.21}$$

hence,

$$\int\int f^*(x, y) \ln f^*(x, y)[f_N(x, y)]^{-1} \, dx \, dy = 0 \tag{2.22}$$

and

$$f^*(x, y) = f_N(x, y) \quad \text{a.e.}, \qquad N \geqq 2n - 1. \tag{2.23}$$

Note that (2.22) implies that the sequence (2.11) reaches a zero value and stays at that value thereafter.

Now let us consider the case when there is no equality in (2.11). Since (2.11) is a monotonic decreasing sequence of nonnegative numbers bounded below it converges to a finite value as $n \to \infty$, hence

$$I(f^*; f_{2n-1}) - I(f^*; f_{2n}) = I(h; h_{2n-1}) \to 0, \tag{2.24}$$

$$I(f^*; f_{2n}) - I(f^*; f_{2n+1}) = I(g; g_{2n}) \to 0. \tag{2.25}$$

It may be shown [Kullback, (1967)] that (2.24) and (2.25) imply

$$\int |h_{2n-1}(y) - h(y)| \, dy \to 0, \qquad \int |g_{2n}(x) - g(x)| \, dx \to 0 \tag{2.26}$$

as $n \to \infty$, or using (2.2),

$$\int |h_N(y) - h(y)| \, dy \to 0, \qquad \int |g_N(x) - g(x)| \, dx \to 0, \qquad N \to \infty. \tag{2.27}$$

Hence, using (2.1), as $n \to \infty$,

$$\begin{aligned} \int\int |f_{2n}(x, y) &- f_{2n-1}(x, y)| \, dx \, dy \\ &= \int\int f_{2n-1}(x, y)|h(y) - h_{2n-1}(y)|[h_{2n-1}(y)]^{-1} \, dx \, dy \\ &= \int |h(y) - h_{2n-1}(y)| \, dy \to 0, \end{aligned} \tag{2.28}$$

$$\begin{aligned} \int\int |f_{2n+1}(x, y) &- f_{2n}(x, y)| \, dx \, dy \\ &= \int\int f_{2n}(x, y)|g(x) - g_{2n}(x)|[g_{2n}(x)]^{-1} \, dx \, dy \\ &= \int |g(x) - g_{2n}(x)| \, dx \to 0, \end{aligned} \tag{2.29}$$

and for any m, as $N \to \infty$,

$$\begin{aligned} \int\int |f_{N+m}(x, y) &- f_N(x, y)| \, dx \, dy \\ &\leqq \int\int |f_{N+m}(x, y) - f_{N+m-1}(x, y)| \, dx \, dy \\ &\quad + \int\int |f_{N+m-1}(x, y) - f_{N+m-2}(x, y)| \, dx \, dy + \cdots \\ &\quad + \int\int |f_{N+1}(x, y) - f_N(x, y)| \, dx \, dy \to 0, \end{aligned} \tag{2.30}$$

so that there exists a function which we write as $f_\infty(x, y)$ defined uniquely a.e.

such that [Titchmarsh, (1939), pp. 386–389]

$$\int\int |f_N(x, y) - f_\infty(x, y)| \, dx \, dy \to 0, \qquad N \to \infty. \tag{2.31}$$

Since, as $N \to \infty$,

$$\int |g(x) - g_\infty(x)| \, dx \leqq \int |g(x) - g_N(x)| \, dx + \int |g_N(x) - g_\infty(x)| \, dx \to 0, \tag{2.32}$$

and

$$\int |h(y) - h_\infty(y)| \, dy \leqq \int |h(y) - h_N(y)| \, dy + \int |h_N(y) - h_\infty(y)| \, dy \to 0; \tag{2.33}$$

it follows that [Titchmarsh, (1939), pp. 386–389]

$$\int |g(x) - g_\infty(x)| \, dx = 0, \qquad g(x) = g_\infty(x) \quad \text{a.e.}, \tag{2.34}$$

$$\int |h(y) - h_\infty(y)| \, dy = 0, \qquad h(y) = h_\infty(y) \quad \text{a.e..} \tag{2.35}$$

It may be shown as before that $f_\infty(x, y) = f^*(x, y)$ a.e..

3. Remarks. It should be clear that the formulation and discussion in terms of bivariate probability densities was a matter of convenience rather than of a limitation imposed by the techniques. To avoid a possible cumbersome notational problem for the general case, results will be stated for four-variate probability densities assuming different marginals given, and which illustrate the general results. Although the same symbol will be used for the probability density and its marginals, this does not imply a common functional form and a single integral sign will be used for multiple integrals.

Let $\pi(x_1, x_2, x_3, x_4)$ be some four-variate probability density and required the four-variate probability density $f(x_1, x_2, x_3, x_4)$ with given marginal probability densities

$$f(x_1), f(x_2), f(x_3), f(x_4), \tag{3.1}$$

such that

$$I(f; \pi) = \int f(x_1, x_2, x_3, x_4) \ln f(x_1, x_2, x_3, x_4)[\pi(x_1, x_2, x_3, x_4)]^{-1} \cdot dx_1 \, dx_2 \, dx_3 \, dx_4 \tag{3.2}$$

is a minimum for all four-variate probability densities with the given marginals. The minimum is attained for

$$\begin{aligned} f_1^*(x_1, x_2, x_3, x_4) &= r(x_1)s(x_2)t(x_3)u(x_4)\pi(x_1, x_2, x_3, x_4), \\ f(x_1) &= r(x_1) \int s(x_2)t(x_3)u(x_4)\pi(x_1, x_2, x_3, x_4) \, dx_2 \, dx_3 \, dx_4, \\ &\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ f(x_4) &= u(x_4) \int r(x_1)s(x_2)t(x_3)\pi(x_1, x_2, x_3, x_4) \, dx_1 \, dx_2 \, dx_3. \end{aligned} \tag{3.3}$$

The iterative solution of the system (3.3) cycles through

$$(3.4)\quad \begin{aligned} f_{4n+1}(x_1, x_2, x_3, x_4) &= f(x_1)[f_{4n}(x_1)]^{-1}f_{4n}(x_1, x_2, x_3, x_4), \cdots, \\ f_{4n+4}(x_1, x_2, x_3, x_4) &= f(x_4)[f_{4n+3}(x_4)]^{-1}f_{4n+3}(x_1, x_2, x_3, x_4), \end{aligned}$$

and the proof of the convergence follows as in the bivariate case.

If the given marginals are

$$(3.5)\quad f(x_1, x_2),\ f(x_1, x_3),\ f(x_1, x_4),\ f(x_2, x_3),\ f(x_2, x_4),\ f(x_3, x_4),$$

then

$$(3.6)\quad \begin{aligned} &f_2^*(x_1, x_2, x_3, x_4) \\ &\qquad = r(x_1, x_2)s(x_1, x_3)t(x_1, x_4)u(x_2, x_3) \\ &\qquad\quad \cdot v(x_2, x_4)w(x_3, x_4)\pi(x_1, x_2, x_3, x_4), \\ &f(x_1, x_2) = r(x_1, x_2) \int s(x_1, x_3)t(x_1, x_4)u(x_2, x_3) \\ &\qquad\quad \cdot v(x_2, x_4)w(x_3, x_4)\pi(x_1, x_2, x_3, x_4)\, dx_3\, dx_4, \\ &\qquad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ &f(x_3, x_4) = w(x_3, x_4) \int r(x_1, x_2) \cdots v(x_2, x_4)\pi(x_1, x_2, x_3, x_4)\, dx_1\, dx_2. \end{aligned}$$

The iterative solution of the system (3.6) cycles through

$$(3.7)\quad \begin{aligned} f_{6n+1}(x_1, x_2, x_3, x_4) &= f(x_1, x_2)[f_{6n}(x_1, x_2)]^{-1}f_{6n}(x_1, x_2, x_3, x_4), \cdots \\ f_{6n+6}(x_1, x_2, x_3, x_4) &= f(x_3, x_4)[f_{6n+5}(x_3, x_4)]^{-1}f_{6n+5}(x_1, x_2, x_3, x_4). \end{aligned}$$

If the given marginals are

$$(3.8)\quad f(x_1, x_2, x_3), f(x_1, x_2, x_4), f(x_1, x_3, x_4), f(x_2, x_3, x_4)$$

then

$$(3.9)\quad \begin{aligned} f_3^*(x_1, x_2, x_3, x_4) &= r(x_1, x_2, x_3)s(x_1, x_2, x_4)t(x_1, x_3, x_4) \\ &\quad \cdot u(x_2, x_3, x_4)\pi(x_1, x_2, x_3, x_4), \\ f(x_1, x_2, x_3) &= r(x_1, x_2, x_3) \int s(x_1, x_2, x_4)t(x_1, x_3, x_4) \\ &\quad \cdot u(x_2, x_3, x_4)\pi(x_1, x_2, x_3, x_4)\, dx_4, \\ &\quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ f(x_2, x_3, x_4) &= u(x_2, x_3, x_4) \int r(x_1, x_2, x_3)s(x_1, x_2, x_4) \\ &\quad \cdot t(x_1, x_3, x_4)\pi(x_1, x_2, x_3, x_4)\, dx_1. \end{aligned}$$

The iterative solution of the system (3.9) cycles through

$$
(3.10) \quad \begin{aligned} f_{4n+1}(x_1, x_2, x_3, x_4) &= f(x_1, x_2, x_3)[f_{4n}(x_1, x_2, x_3)]^{-1} f_{4n}(x_1, x_2, x_3, x_4), \\ &\cdots, \\ f_{4n+4}(x_1, x_2, x_3, x_4) &= f(x_2, x_3, x_4)[f_{4n+3}(x_2, x_3, x_4)]^{-1} \\ &\quad \cdot f_{4n+3}(x_1, x_2, x_3, x_4). \end{aligned}
$$

The results for other possible combinations of given marginals are left to the reader.

Since (3.8) $\Rightarrow$ (3.5) $\Rightarrow$ (3.1), it is clear that

$$
(3.11) \qquad I(f_3^*; \pi) \geqq I(f_2^*; \pi) \geqq I(f_1^*; \pi) \geqq 0.
$$

Applying the relation (1.10), it follows that

$$
(3.12) \quad \begin{aligned} I(f_3^*; \pi) &= I(f_2^*; \pi) + I(f_3^*; f_2^*), \\ I(f_2^*; \pi) &= I(f_1^*; \pi) + I(f_2^*; f_1^*), \\ I(f_3^*; \pi) &= I(f_1^*; \pi) + I(f_3^*; f_1^*), \\ I(f_3^*; f_1^*) &= I(f_3^*; f_2^*) + I(f_2^*; f_1^*). \end{aligned}
$$

If in (1.7) $\pi(x, y) = \pi(x)\pi(y)$ and in (3.3) $\pi(x_1, x_2, x_3, x_4) = \pi(x_1)\pi(x_2) \cdot \pi(x_3)\pi(x_4)$, then it readily follows that in (1.7) $f^*(x, y) = g(x)h(y)$ and in (3.3) $f_1^*(x_1, x_2, x_3, x_4) = f(x_1)f(x_2)f(x_3)f(x_4)$, that is, if the π density is a product of its marginals (independence) the value of the f_1^* density does not depend on the π density and is the product of the given marginals.

REFERENCES

[1] Good, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* **34** 911–934.

[2] Good, I. J. (1966). How to estimate probabilities. *J. Inst. Math. Applics.* **2** 364–383.

[3] Ireland, C. T. and Kullback, S. (1968). Contingency tables with given marginals. *Biometrika* **55** 315–324.

[4] Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.

[5] Kullback, S. (1967). A lower bound for discrimination information in terms of variation. I.E.E.E. Transactions on *Information Theory*. ***IT*-13** 126–127.

[6] Kullback, S. and Khairat, M. A. (1966). A note on minimum discrimination information. *Ann. Math. Statist.* **37** 279–280.

[7] Rényi, A. (1962). *Wahrscheinlichkeitsrechnung*. VEB Deutscher Verlag der Wissenschaften, Berlin.

[8] Titchmarsh, E. C. (1939). *The Theory of Functions*, 2nd Ed. Oxford Univ. Press.