The Annals of Mathematical Statistics
1968, Vol. 39, No. 4, 1186–1190

THE GEOMETRY OF AN $r \times c$ CONTINGENCY TABLE[1]

BY STEPHEN E. FIENBERG

Harvard University

1. Introduction. Any contingency table can be normalized to have entries which add to one, and then all possible $r \times c$ two-way tables can be represented by points within the $(rc - 1)$-dimensional simplex

$$S_{rc} = \{(x_{11}, x_{12}, \cdots, x_{1c}; \cdots; x_{r1}, \cdots, x_{rc}) : x_{ij} \geqq 0, \textstyle\sum_{i,j} x_{ij} = 1\} \tag{1.1}$$

in rc-space. A deeper understanding of the geometry associated with this simplex might allow us to deal with the corresponding contingency tables in a more enlightened manner.

In a previous paper, [2], ideas about 2×2 contingency tables were discussed in terms of the geometry of the 3-dimensional simplex. Here we generalize these ideas and in particular we derive the loci of (a) all points corresponding to tables whose rows and columns are independent, (b) all points corresponding to tables with a given interaction structure, and (c) all points corresponding to a table with a fixed set of marginals. Finally we conclude with a discussion of the generalization of our results to multidimensional tables.

2. The simplex of reference. We examine the simplex S_{rc} by means of rc-dimensional barycentric co-ordinates, [1], and choose the simplex of reference (with vertices A_{ij} for $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, c$) so that

$$\begin{aligned} A_{11} &= (1, 0, 0, \cdots, 0, 0, 0), \\ A_{12} &= (0, 1, 0, \cdots, 0, 0, 0), \\ &\ \ \vdots \\ A_{r(c-1)} &= (0, 0, 0, \cdots, 0, 1, 0) \\ A_{rc} &= (0, 0, 0, \cdots, 0, 0, 1) \end{aligned} \tag{2.1}$$

correspond respectively to the $r \times c$ tables

$$\begin{matrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0, \end{matrix} \qquad \begin{matrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0, \end{matrix}$$

Received 25 September 1967.

[1] This work was facilitated in part by a grant from the National Science Foundation (GS-341), and in part by a doctoral fellowship from the Canada Council.

(2.2) through to

$$\begin{matrix} 0 & 0 & \cdots & 0 & 0 & & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0, & & 0 & 0 & \cdots & 0 & 1. \end{matrix}$$

The general point

$$P = (p_{11}, p_{12}, \cdots, p_{1c}; \cdots; p_{r1}, \cdots, p_{rc}) \tag{2.3}$$

corresponds to the general $r \times c$ table with cell entries $p_{ij}(\sum_{i,j} p_{ij} = 1)$.

Thus there is a 1–1 correspondence between points in the simplex and population $r \times c$ tables, although for sample $r \times c$ tables the correspondence is with all points which have rational co-ordinates.

3. The manifold of independence. Following Sommerville [5] we refer to a linear space of n dimensions as an *n-flat*. Thus an n-flat is determined by $n + 1$ points and every m-flat $(m < n)$ which is determined by $m + 1$ of these points lies entirely within the n-flat.

In the $(rc - 1)$-dimensional space containing S_{rc} there exist c $(r - 1)$-flats such that each contains r distinct vertices of the simplex and all of the points corresponding to tables with their probability concentrated in one specific column. The vertices contained in the jth such $(r - 1)$-flat are $A_{1j}, A_{2j}, \cdots, A_{rj}$.

Now we fix a general point

$$T_1 = (t_1, 0, \cdots, 0; t_2, 0, \cdots; t_r, 0, \cdots, 0) \tag{3.1}$$

in the first $(r - 1)$-flat, where $\sum_{i=1}^{r} t_i = 1$ and $t_i \geqq 0 \ \forall\ i$. We denote the analogous points (whose corresponding tables have the same row margins) in the remaining $c - 1(r - 1)$-flats by $T_2, T_3, \cdots, T_c$ respectively (for example, $T_2 = (0, t_1, 0, \cdots, 0; \cdots; 0, t_r, 0, \cdots, 0)$). Thus the $(c - 1)$-flat containing $T_1, T_2, \cdots, T_c$, consists of all points

$$I = (t_1s_1, t_1s_2, \cdots, t_1s_c; t_2s_1, \cdots; t_rs_1, t_rs_2, \cdots, t_rs_c). \tag{3.2}$$

Its intersection with S_{rc} consists of all such points with $\sum_{j=1}^{c} s_j = 1$ and $s_j \geqq 0 \ \forall\ j$. But these points correspond to tables which are said to be independent. Now by allowing the t_i to vary subject to the constraints $\sum_{i=1}^{r} t_i = 1$ and $t_i \geqq 0 \ \forall\ i$, we get a family of nonintersecting $(c - 1)$-flats, which contain all points corresponding to independent tables. These $(c - 1)$-flats generate what we will call the *manifold of independence.*

Alternatively, we might have set out by discussing the existence of r $(c - 1)$-flats such that each contains c distinct vertices of S_{rc} and all of the points corresponding to tables with their probability concentrated in one specific row. The vertices joined by the ith such $(c - 1)$-flat are $A_{i1}, A_{i2}, \cdots, A_{ic}$. Continuing as before we find that the *manifold of independence* is also generated by a family of nonintersecting $(r - 1)$-flats.

The tables corresponding to points on any one of the family of nonintersecting $(c - 1)$-flats have the same row margins (totals) while the tables corresponding to points on any one of the family of nonintersecting $(r - 1)$-flats have the same column margins. Note that each $(c - 1)$-flat meets each $(r - 1)$-flat in a single point, corresponding to the independent $r \times c$ table with margins defined by the $(c - 1)$-flat and the $(r - 1)$-flat.

A manifold in n-space is said to have co-dimension d iff there exists a 1–1 onto mapping taking the manifold into an $(n - d)$-flat with nonzero $(n - d)$ dimensional volume. Thus the *manifold of independence* has co-dimension $(r - 1)(c - 1)$. For 2×2 tables [2] the *manifold of independence*, which has co-dimension 1, is a hyperbolic paraboloid, and the families of nonintersecting $(c - 1)$-flats and $(r - 1)$-flats are simply two families of straight lines or "rulings".

4. The manifold of constant interaction. Our method of constructing the *manifold of independence* is equivalent to saying that a general point P lies on the manifold iff all of the $(r - 1)(c - 1)$ crossproducts

$$\alpha_{ij} = p_{ij}p_{(i+1)(j+1)}/p_{(i+1)j}p_{i(j+1)} \qquad \text{for } i = 1, 2, \cdots, r - 1; \quad j = 1, 2, \cdots, c - 1; \tag{4.1}$$

are equal to 1. When the α_{ij} are not all equal 1, we say that the corresponding table expresses interaction or nonindependence. Such a use of the α_{ij} has been examined by, among others, Goodman [3] and Lindley [4]. Note that $\infty \geqq \alpha_{ij} \geqq 0 \;\forall\; i, j$.

We can now construct *manifolds of constant interaction* by a procedure similar to that used in Section 3 to define the *manifold of independence*. First we fix a set of $\overset{0}{\alpha}_{ij}$ such that $\overset{0}{\alpha}_{ij} \not\equiv 1 \;\forall\; i, j$ and no $\overset{0}{\alpha}_{ij} = 0$ or ∞. Then we look at the c $(r - 1)$-flats, $A_{1j}A_{2j} \cdots A_{rj}$, for $j = 1, 2, \cdots, c$. Again we fix the general point T_1 (3.1) in the first $(r - 1)$-flat, but this time we choose the analogous points (denoted by T_j^* for $j = 2, 3, \cdots, c$) in the remaining $c - 1$ $(r - 1)$-flats, so that the $[(i - 1)c + j]$th co-ordinate of T_j^* is

$$V_j t_i \cdot \left(\textstyle\prod_{k=0}^{i-1} \prod_{l=1}^{j-1} \overset{0}{\alpha}_{kl}\right) \qquad \text{for } i = 1, 2, \cdots, r, \tag{4.2}$$

where $\overset{0}{\alpha}_{0l} = 1 \;\forall\; l$, and

$$1/V_j = \textstyle\sum_{i=1}^{r} t_i\left(\prod_{k=0}^{i-1} \prod_{l=1}^{j-1} \overset{0}{\alpha}_{kl}\right). \tag{4.3}$$

Then the $(c - 1)$-flat containing $T_1, T_2^*, T_3^*, \cdots, T_c^*$, has an intersection with the simplex, S_{rc}, consisting of points which correspond to tables with the fixed crossproduct values $\overset{0}{\alpha}_{ij}$. By letting the t_i vary subject to the constraints $\sum_i t_i = 1$ and $t_i > 0 \;\forall\; i$, we get a family of nonintersecting $(c - 1)$-flats which contain all points corresponding to tables with the fixed value of $\overset{0}{\alpha}_{ij}$. We call the manifold generated by this family of $(c - 1)$-flats, the *manifold of constant interaction* $(\overset{0}{\alpha}_{ij})$.

Alternatively, we might begin with the r $(c - 1)$-flats, $A_{i1}A_{i2}, \cdots, A_{ic}$, for $i = 1, 2, \cdots, r$. Then we find that the *manifold of constant interaction* (α_{ij}^0) is also generated by a family of nonintersecting $(r - 1)$-flats.

The tables corresponding to points on any one of the family of $(c - 1)$-flats or on any one of the family of $(r - 1)$-flats do not have a common set of row or column margins, as is the case for the generating flats on the *manifold of independence*. Again we note that each of the $(c - 1)$-flats meets each of the $(r - 1)$-flats in exactly one point.

The *manifold of constant interaction* (α_{ij}^0) also has co-dimension $(r - 1)(c - 1)$. For the 2×2 table the co-dimension is thus 1, and we can show that the manifold is a hyperboloid of one sheet, and the generating flats are simply two families of straight lines [2]. When $r > 2$ or $c > 2$ the manifold is simply the intersection of $(r - 1)(c - 1)$ quadric manifolds corresponding to the $(r - 1)(c - 1)$ cross-products α_{ij}^0.

When some of the α_{ij}^0 are equal to 0 or to ∞ the *manifold of constant interaction* (α_{ij}^0) becomes degenerate. For 2×2 tables there is only one crossproduct, α_{11}^0, and thus there are only two degenerate manifolds. Each turns out to be a pair of faces of the tetrahedron of reference which meet in an edge of the tetrahedron not on the surface of independence. For $r > 2$ or $c > 2$ the degeneracies are more involved, and we will not discuss them in any detail.

5. Tables with fixed margins. Now let us take a general point P (2.3) and a point I (3.2) on the *manifold of independence* (both within the simplex). The direction numbers of the line PI are given by

$$(p_{11} - t_1s_1, p_{12} - t_1s_2, \cdots, p_{rc} - t_rs_c). \tag{5.1}$$

It is a simple exercise to show that requiring PI to be orthogonal to the $(r - 1)$ flat and the $(c - 1)$-flat on the *manifold of independence* which pass through the point with all its co-ordinates equal to $1/rc$, is equivalent to the $r + c$ linear constraints

$$\textstyle\sum_{i=1}^{r} p_{ij} = s_i \quad \text{for} \quad j = 1, 2, \cdots, c, \tag{5.2}$$

and

$$\textstyle\sum_{j=1}^{c} p_{ij} = t_i \quad \text{for} \quad i = 1, 2, \cdots, r. \tag{5.3}$$

These reduce to $r + c - 2$ constraints since we already know that $\sum_{ij} p_{ij} = 1$, $\sum_j s_i = 1$, and $\sum_i t_i = 1$. Thus the locus of all points corresponding to tables with fixed margins (s_j for $j = 1, 2, \cdots, c$, and t_i for $i = 1, 2, \cdots, r$) is the intersection of the simplex with the $(r - 1)(c - 1)$-flat orthogonal to the $(r - 1)$-flat and the $(c - 1)$-flat on the *manifold of independence* which pass through the point with equal co-ordinates. This $(r - 1)(c - 1)$-flat intersects each *manifold of constant interaction* (provided no $\alpha_{ij} = 0$ or ∞) in exactly one point, and, as we have already seen in Section 3, it intersects the *manifold of independence* in exactly one point.

By first choosing two sets of margins (row and column) and fixing a point in the corresponding $(r-1)(c-1)$-flat we see geometrically that an $r \times c$ table is uniquely determined by its margins and the $(r-1)(c-1)$ crossproducts (4.1).

6. Multidimensional tables. We can easily extend some of our geometrical ideas to multidimensional tables. Here we will briefly look at three-way $r \times c \times d$ contingency tables which can be represented by points within the $(rcd - 1)$-dimensional simplex with rcd vertices.

We can easily show, by extending the arguments of Section 3, that the locus of all points corresponding to independent tables (i.e. those tables with no second or third order interaction) is a manifold generated by *three* families of nonintersecting flats, of dimensions $r-1$, $c-1$, and $d-1$. This manifold has a co-dimension $[(r-1)(c-1)(d-1) + (r-1)(d-1) + (r-1)(c-1) + (c-1)(d-1)]$. Also, the tables with constant margins correspond to points on a flat of dimension $[(r-1)(c-1)(d-1) + (r-1)(d-1) + (r-1)(c-1) + (c-1)(d-1)]$ which is orthogonal to the generating flats on this *manifold of independence* which pass through the point with co-ordinates all equal to $1/rcd$. Each flat of constant margins meets the *manifold of independence* in exactly one point.

We conjecture the existence of a manifold of co-dimension $(r-1)(c-1) \cdot (d-1)$, which contains all points corresponding to tables with no third order interaction, and which contains the *manifold of independence* as a submanifold.

7. Acknowledgment. The author wishes to thank John P. Gilbert, Paul Holland, and Frederick Mosteller for helpful discussions which aided in the preparation of this paper.

REFERENCES

[1] Coxeter, H. S. M. (1961). *Introduction to Geometry.* Wiley, New York.

[2] Fienberg, S. E., and Gilbert, J. P. (1967). The geometry of a two by two contingency table. Unpublished manuscript.

[3] Goodman, L. A. (1964). Simultaneous confidence limits for crossproduct ratios in contingency tables. *J. Roy. Statist. Soc. Ser. B.* **26** 86–102.

[4] Lindley, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.* **35** 1622–1643.

[5] Sommerville, D. M. Y. (1958). *An Introduction to the Geometry of N Dimensions.* Dover, New York.