Conceptual and Visual Models for Categorical Data

Michael FRIENDLY

A dynamic conceptual model for categorical data is described that likens observations to gas molecules in a pressure chamber. In this physical model frequency corresponds to pressure, and fitting a statistical model by maximum likelihood corresponds to minimizing energy or balancing of forces. The model provides neat explanations of many results for categorical data, extends readily to multiway tables, and provides a rationale for the graphic representation of counts by area or visual density.

KEY WORDS: Graphical display; Iterative proportional fitting; Maximum likelihood; Mechanical models; Mosaic display; Newton–Raphson iteration.

1. CONCEPTUAL MODELS AND VISUAL METAPHORS

Visual representation of data depends fundamentally on an appropriate visual scheme for mapping numbers into graphic patterns (Bertin 1983). One reason for the widespread use of graphical methods for quantitative data is the availability of a natural visual mapping: magnitude can be represented by length, as in a bar chart, or by position along a scale, as in dot charts and scatterplots. One reason for the relative paucity of graphical methods for categorical data may be that a natural visual mapping for frequency data is not so apparent. This article describes and extends a conceptual model for categorical data due to Sall (1991b). This model suggests that natural visual attributes for categorical data are to depict frequency by area or observation density.

Closely associated with the idea of a visual metaphor is a conceptual model that helps you interpret what is shown in a graph. A good conceptual model for a graphical display will have deeper connections with underlying statistical ideas as well. For quantitative data, position along a scale can be related to mechanical models in which fitting data by least squares or least absolute deviations correspond directly to balancing forces or minimizing potential energy (Farebrother 1987). The mechanical model for least squares regression, for example, likens each observation to a unit mass connected vertically to a rod by springs of unit modulus. Sall (1991a) shows how this mechanical model neatly describes the effects of sample size on power of a test, the leverage of outlying observations in regression, principal components, and collinearity among others.

2. CATEGORICAL DATA

With categorical data each response is classified into a distinct category. Imagine we are observing the hair color and eye color of students in a large class, as in the data [from Snee (1974)] shown in Table 1. What sort of conceptual model can we use to represent the frequencies of these categories? We first consider the one-way case, for the marginal frequencies of hair color.

2.1 Urn Model for Multinomial Sampling

If the total sample size, n, is fixed (multinomial sampling), maximum likelihood estimation of the category probabilities π_i leads to maximizing the sum of logs of the probabilities attributed to the events that actually occurred, or equivalently, choosing the values $\widehat{\pi}_i$ of π_i which minimize $-\sum_i n_i \log \pi_i$. One model would be a set of four urns labeled with the hair colors and containing marbles representing each observation (Fig. 1).

This model serves as a metaphor for the familiar bar chart display in which each count n_i is portrayed by the area filled by observations. When the urns are of equal width, count is also reflected by height, but in the general case, count is proportional to area. However, the urn model is a static one and provides no further insights. It does not relate to the concept of likelihood or to the constraint that the probabilities sum to 1.

2.2 Dynamic Model: Pressure and Energy

We can change the conceptual model to a dynamic, physical one by giving each observation a force (Fig. 2). One way to do that is to consider the observations as molecules of an ideal gas confined to a cylinder whose volume can be varied with a movable piston (Sall 1991b). If *everyone* had red hair, observing someone with red hair would convey no information. So, this mechanical model is set up so that a probability of 1.0 corresponds to ambient pressure, whence the force exerted on the piston is 0. An actual probability of red hair equal to p means that the same number of observations is squeezed down to a chamber of height p. By Boyle's law (that pressure times volume is constant) the pressure is proportional to $1/p$. In the figure, pressure is shown by *observation density*, the number of observations per unit area. Hence, the graphical analog of pressure is that a count can be represented visually by observation density when the count is fixed and area is varied (or by area when the observation density is fixed as in Figure 1).

Now, the work done on the gas (or potential energy imparted to it) by compression of a small distance δy is the force on the piston times δy, which equals the pressure times the change in volume. Hence, the potential energy of a gas at height p is $\int_p^1 (1/y)\, dy$, which is $-\log(p)$, so the energy in this model corresponds to negative log-likelihood.

2.2.1 Fitting Probabilities: Minimum Energy, Balanced Forces. Maximum likelihood estimation means literally finding the values, $\widehat{\pi}_i$, of the parameters under

Michael Friendly is Associate Professor of Psychology and Director of the Statistical Consulting Service, York University, Toronto, ON, Canada M3J 1P3. e-mail: FRIENDLY@VM1.YorkU.CA. John Sall made numerous contributions to this paper. Thanks are also due to William Farebrother, Paul Herzberg, Georges Monette, and the referees for helpful comments.

© 1995 American Statistical Association

Table 1. Hair-Color Eye-Color Data

Eye Color	Hair Color: Black	Brown	Red	Blond	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

which the observed data would have the highest probability of occurrence. We take derivatives of the (log-) likelihood function with respect to the parameters, set these to zero, and solve. For a multiparameter model for quantitative data this often leads to simultaneous equations that must be solved by inverting a matrix, but it has always struck me as curious that, in many cases, maximum likelihood estimation for categorical data simply sets parameter values equal to corresponding sample quantities:

$$\frac{\partial \log L}{\partial \pi_i} = 0 \Rightarrow \frac{n_1}{\pi_1} = \frac{n_2}{\pi_2} = \cdots = \frac{n_c}{\pi_c} \Rightarrow \hat{\pi}_i = \frac{n_i}{n} = p_i.$$

In the mechanical model (Fig. 3) this corresponds to stacking the gas containers with movable partitions between them, with one end of the bottom and top containers fixed at 0 and 1. The observations exert pressure on the partitions, the likelihood equations are precisely the conditions for the forces to balance, and the partitions move so that each chamber is of size $p_i = n_i/n$. Each chamber has potential energy of $-\log p_i$, and the total energy, $-\sum_i^c n_i \log p_i$ is minimized. The constrained top and bottom force the probability estimates to sum to 1, and the number of movable partitions is literally and statistically the degrees of freedom of the system. (Figure 3, like all other figures in this paper, is graphically exact: the area of each chamber is proportional to the marginal frequencies of hair color, and is filled with the corresponding number of points.)

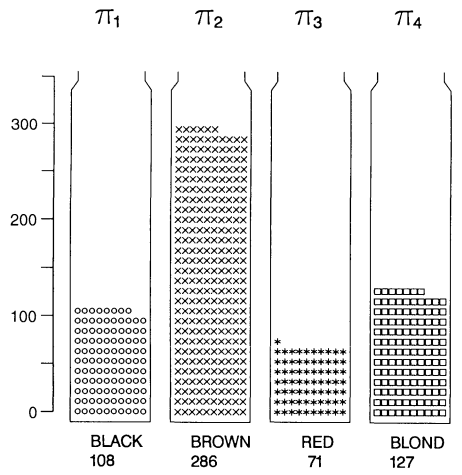

Figure 1. Urn Model for Multinomial Sampling. Each observation in Table 1 is represented by a token classified by hair color into the appropriate urn. This model provides a basis for the bar chart, but does not yield any further insights.

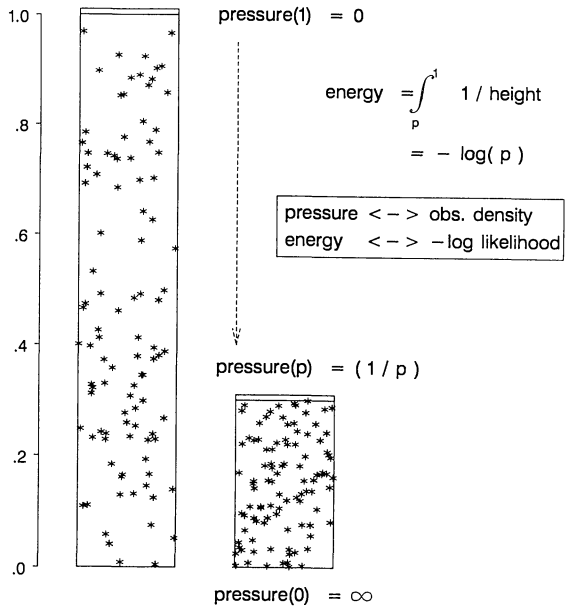

Figure 2. Dynamic Conceptual Model for Categorical Data. Frequency of observations corresponds to pressure of gas in a chamber, shown visually as observation density; negative log-likelihood corresponds to the energy required to compress the gas to a height p.

The general principle is that fitting a statistical model by minimizing the negative likelihood corresponds to a state of minimum energy subject to a set of analogous constraints; the first-order (derivatives = 0) conditions for the solution are reflected in a state of equilibrium or balancing of forces. These physical and statistical properties are shown visually as equal observation densities across the cells in Figure 3. See Farebrother (1988) for another description of this idea and its application to multidimensional scaling. A more formal statement is implicit in a

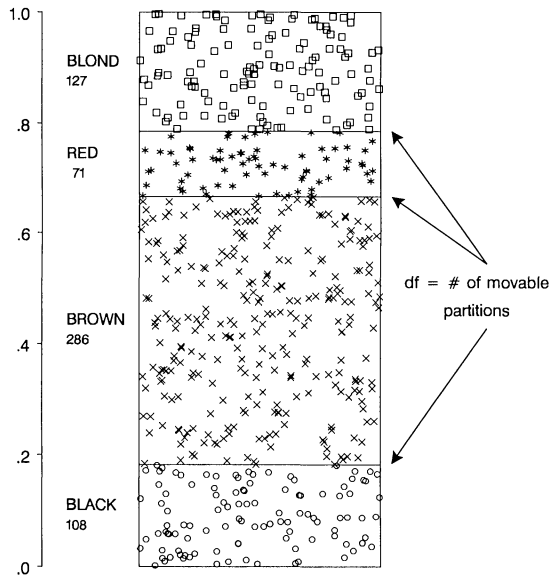

Figure 3. Fitting Probabilities for a One-Way Table. The movable partitions naturally adjust to positions of balanced forces, which is the minimum energy configuration.

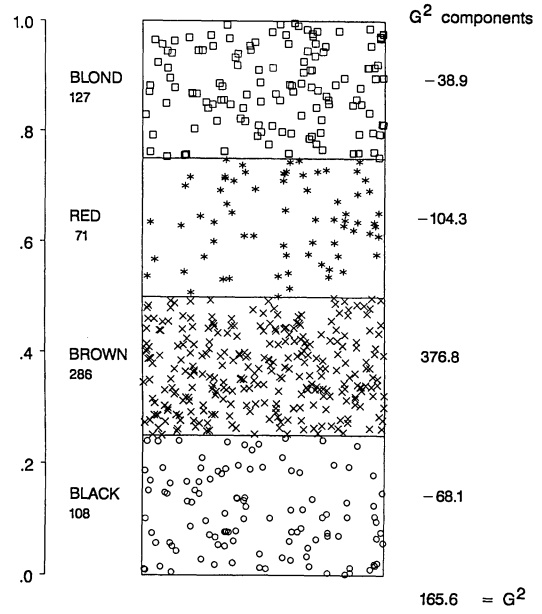

Figure 4. Testing a Hypothesis. The likelihood ratio G^2 measures how much energy is required to move the partitions to constrain the data to hypothesized probabilities. The components of G^2 indicate the degree to which each chamber has low or high pressure, relative to the balanced state.

duality theorem of Good (1963) that gives a relationship between maximum likelihood estimation and maximum entropy.

2.2.2 Testing a Hypothesis: How Much More Energy? This mechanical model also explains how we test hypotheses about the true probabilities (Fig. 4). To test the hypothesis that the four hair color categories are equally probable,

$$H_0: \pi_1 = \pi_2 = \pi_3 = \pi_4 = \tfrac{1}{4}$$

simply force the partitions to move to the hypothesized values and measure how much energy is required to force the constraint. Some of the chambers will then exert more pressure, some less than when the forces are allowed to balance without these additional restraints. The change in energy in each compartment is then $-(\log p_i - \log \pi_i) = -\log(p_i/\pi_i)$, the change in negative log-likelihood. Sum these up and multiply by 2 to get the likelihood ratio G^2. This gives a concrete instantiation of the interpretation of G^2 as a measure of the effort to maintain belief in the hypothesis in the face of the data.

2.2.3 Other Neat Explanations. The pressure model provides simple explanations of other results and phenomena, some of which are paradoxical or difficult to appreciate when demonstrated formally.

- For example, it may seem peculiar that components of the likelihood ratio G^2 are both positive and negative, whereas other X^2 quantities are sums of squares. From the dynamic model shown in Figure 4, we see that negative contributions to G^2 correspond to relative low-pressure cells with fewer observations than hypothesized and positive contributions to relative high-pressure cells.
- The effects of sample size on power of the test is also made apparent: More observations means more pressure in each compartment, so it takes more energy to move the partitions and the test is sensitive to smaller differences between observed and hypothesized probabilities. When the p_i are fixed, the force in each chamber and the observed G^2 increase linearly with total sample size.
- Finally, the dynamic model explains the curious fact that although other sampling schemes such as Poisson sampling ascribe quite different probability distributions to the observed counts, they usually lead to identical maximum likelihood estimates: The sampling scheme describes how the observations enter the chambers; once they are there, the forces and energy determine the outcome.

3. MULTIWAY TABLES

The dynamic pressure model extends readily to multiway tables. For a two-way table of hair color and eye color, partition the sample space according to the marginal proportions of eye color, and then partition the observations for each eye color according to hair color as before (Fig. 5). Within each column the forces balance as before, so that the height of each chamber is n_{ij}/n_{i+}. Then the area of each cell is proportional to the MLE of the cell probabilities,

$$\left(\frac{n_{i+}}{n}\right)\left(\frac{n_{ij}}{n_{i+}}\right) = \frac{n_{ij}}{n} = p_{ij},$$

which again is the sample cell proportion.

For a three-way table, the natural physical model is a cube with its third dimension partitioned according to conditional frequencies of the third variable, given the first two. The volumes of each chamber balance when the pressures balance. Since graphic representations are

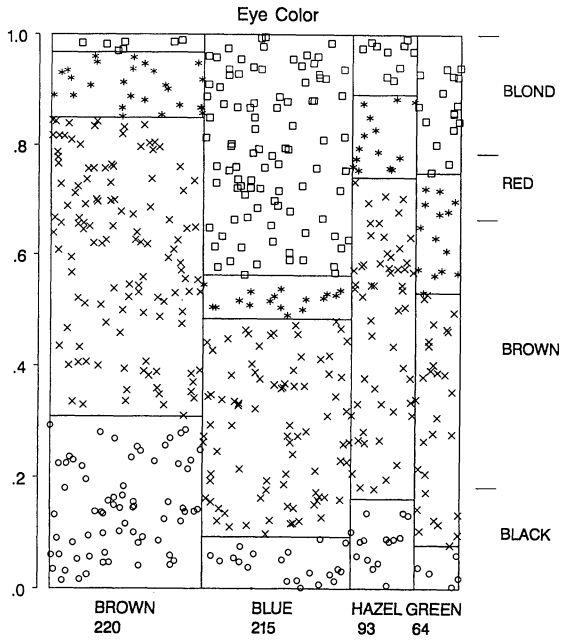

Figure 5. Two-Way Tables. For multiple samples, the model represents each sample by a stack of pressure chambers whose width is proportional to the marginal frequencies of one variable.

difficult in more than two dimensions, we can extend the model to higher dimensions (with some loss of fidelity between the physical and visual models) by partitioning a two-dimensional graph, alternating between vertical and horizontal divisions for each new variable. The graphic representation of this conceptual model is, in fact, the same as the mosaic display of Hartigan and Kleiner (1981) as extended by Friendly (1991, 1993).

3.1 Testing Models of Independence

For a two-way table of size $I \times J$ the hypothesis of independence is formally the same as the hypothesis that conditional probabilities (of hair color) are the same in all strata (eye colors). To test this hypothesis, simply force the partitions to align and measure the total additional energy required to effect the change (Fig. 6). The degrees of freedom for the test is again the number of movable partitions, $(I-1)(J-1)$. More precisely, it is the difference in the number of movable partitions between the case of the unconstrained model ($(J-1)+(I-1)J = IJ-1$) and the hypothesis-constrained model ($(I-1)+(J-1)$). Alternatively, starting with the constrained configuration in Figure 6, the degrees of freedom is the number of cuts you must make in the horizontal partitions to let them float back to the unconstrained case in Figure 5.

For a 2×2 table (or a 2×2 portion of a larger table) the row and column variables are independent when the ratio of pressures in one row (or column) equals the ratio of pressures in the other row (column). So we can also measure departure from independence by how far the ratio of these ratios, $(n_{11}/n_{12})/(n_{21}/n_{22})$, departs from 1. This is the familiar sample odds ratio.

For three-way tables there are various models of independence. Complete independence, the log-linear model $[A]\,[B]\,[C]$, corresponds to the cube in which all chambers

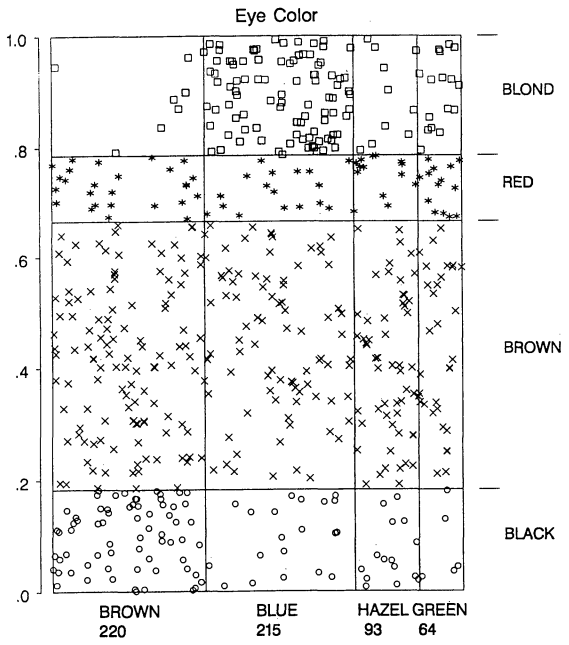

Figure 6. Testing Independence. The chambers are forced to align with both sets of marginal frequencies, and the likelihood ratio G^2 again measures the additional energy required.

are forced to conform to the one-way marginals,

$$\pi_{ijk} = \pi_{i++}\,\pi_{+j+}\,\pi_{++k}$$

for all i, j, k. G^2 is again the total additional energy required to move the partitions from their positions in the saturated model in which the volume of each cell is $p_{ijk} = n_{ijk}/n$ (so the pressures balance) to the positions where each cell is a cube of size $\pi_{i++} \times \pi_{+j+} \times \pi_{++k}$. Other hypotheses of independence have a similar representation in the pressure model. For example, in the model of joint independence, $[AB]\,[C]$, we have $\pi_{ijk} = \pi_{ij+}\,\pi_{++k}$, so the partitions are first moved to match the marginal proportions p_{ij+} of the first two variables jointly. Then the partitions of the third dimension are moved so that in each ij plane they all align with the marginal proportions p_{++k} of the third variable.

In general, the marginals which are the minimal sufficient statistics for the model under test are those imposed on the data. This is exactly what happens in the maximum likelihood solutions for log-linear models. Let $\mathbf{n}$ and $\mathbf{m}$ denote column vectors of observed and expected frequencies, $\{n_{ijk\ldots}\}$ and $\{m_{ijk\ldots}\}$. Casting the model in the form $\log \mathbf{m} = \mathbf{X}\beta$, where $\mathbf{X}$ is the design matrix of the hypothesis, and β is the vector of parameters, the solution,

$$\mathbf{X}'\mathbf{n} = \mathbf{X}'\widehat{\mathbf{m}} \tag{1}$$

equates sufficient marginals of the data to their expected frequencies under the model.

3.2 Iterative Proportional Fitting

For three-way (and higher) tables some log-linear models have direct estimates of expected cell frequencies. The cases where direct estimates exist are analogous to the two-way case, where the estimates under the hypothesized model are products of the sufficient marginals. Here we see that the partitions in the observation space can be moved directly in planar slices to their positions under the hypothesis, so that iteration is unnecessary.

When direct estimates do not exist, the MLE's can be estimated by iterative proportional fitting (IPF). This process simply matches the partitions corresponding to each of the sufficient marginals of the fitted frequencies to the same marginals of the data. For example, for the log-linear model $[AB]\,[BC]\,[AC]$, the sufficient statistics are $\{n_{ij+}\}$, $\{n_{i+k}\}$, and $\{n_{+jk}\}$. The conditions that the fitted margins must equal these observed margins from (1) are

$$\frac{n_{ij+}}{\widehat{m}_{ij+}} = \frac{n_{i+k}}{\widehat{m}_{i+k}} = \frac{n_{+jk}}{\widehat{m}_{+jk}} = 1, \tag{2}$$

which is equivalent to balancing the forces in each fitted marginal. The steps in IPF follow directly from equation (2). For example, the first step in cycle $t+1$ of IPF matches the frequencies in the $[AB]$ marginal table,

$$\widehat{m}_{ijk}^{(t+1)} = \widehat{m}_{ijk}^{(t)} \left(\frac{n_{ij+}}{\widehat{m}_{ij+}^{(t)}} \right), \tag{3}$$

which makes the forces balance when (3) is summed over variable C: $\widehat{m}_{ij+}^{(t+1)} = \widehat{n}_{ij+}$. The other steps in each

cycle make the forces balance in the $[BC]$ and $[AC]$ margins.

3.3 Visual Demonstration of IPF

Figure 7 shows a format (suggested by Sall, personal communication) in which the data, fitted frequencies, and marginal frequencies can all be displayed for a three-way table (Table 2, contrived data). The three-way table is first restructured as a two-way table in which one variable forms the columns and the combinations of the remaining two variables form the rows. In Figure 7, for example, the data have been reshaped as an $I \times JK$ table.

The fitted frequencies for this restructured table are used to partition the display. The width of each chamber is proportional to the marginal frequency $\hat{m}_{i++}$ of variable I, and the height is proportional to the conditional probabilities $\hat{m}_{ijk}/\hat{m}_{i++}$. Hence, the areas are proportional to the fitted cell frequencies, $\hat{m}_{ijk}$. The observed frequencies are shown

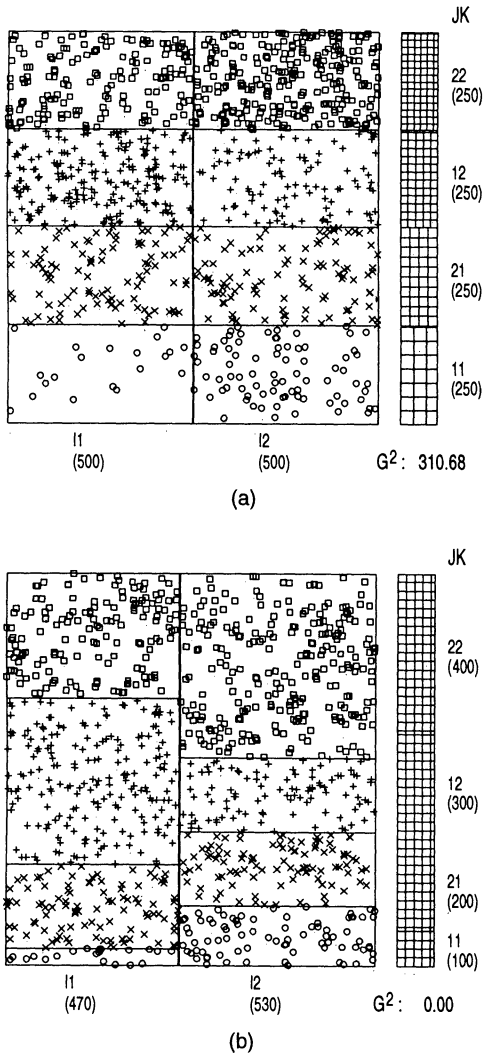

Figure 7. Graphical Display for a Three-Way Table, Showing Observed Counts, and Fitted Cell and Marginal Frequencies. (a) Equal probabilities: I by JK, (b) Saturated model: I by JK. Each cell has an area proportional to fitted frequency and contains the observed n_{ijk} points. The marginal bars at the right have areas proportional to the fitted JK margins, $\hat{m}_{+jk}$ (shown numerically in parentheses) and contain cross-hatched boxes proportional to the observed n_{+jk}.

Table 2. A Three-Way Table (Contrived Data)

	J1		J2	
	I1	I2	I1	I2
K1	20	80	100	100
K2	200	100	150	250

by the number of points in each chamber. Thus, when the fitted and observed frequencies are equal, all cells will appear to be equally densely filled.

The marginal fitted frequencies, $\hat{m}_{+jk}$ for the rows of the restructured table are shown by the heights of the bars to the right and the one-way fitted marginal, m_{++k} is shown by the widths of the columns. (These fitted margins are also shown numerically in parentheses.) These bars are then filled by lines spaced so that the number of boxes within each bar is proportional to the *observed* marginals, n_{+jk}. Hence, when the observed and fitted marginals are balanced, the lines in each bar will be equally spaced and the bars will also appear to be equally densely filled.

Thus, Figure 7a shows the data and fitted frequencies under the equiprobability model where all cells have the same expected frequency, $\hat{m}_{ijk} = n_{+++}/IJK$. The low observed frequency in cell (1, 1, 1) and high frequency in cell (2, 2, 2) appear as chambers with low and high density (pressure). Since the marginal fitted frequencies $\hat{m}_{+jk}$ are not equal to the corresponding observed n_{+jk} margins, the bars vary in density. Figure 7b, however, shows the data and fitted frequencies under the saturated model. Since $\hat{m}_{ijk} = n_{ijk}$, we observe that all the cells have the same density of points; since $\hat{m}_{+jk} = n_{+jk}$, all the marginal bars have the same density of lines.

3.3.1 Balancing Forces. We will show how iterative proportional fitting works in terms of balancing forces in the fitted marginals for the model $[AB]$ $[AC]$ $[BC]$, the only three-way model for which direct estimates do not exist. To do this we need to see all three sets of two-way marginals simultaneously, so Figure 8 shows in the first column the observed counts restructured in the three possible ways. Iterative proportional fitting is initialized (cycle 0) by setting all $\hat{m}_{ijk}^{(0)} = 1$, which is equivalent to the equiprobability model $\hat{m}_{ijk} = n_{+++}/IJK$. The actual counts are shown in cells of size proportional to these initial estimates in the second column of Figure 8. (The panels of Figure 7, b and a, are the same as the panels in the bottom row of Figure 8.)

Now, observe that the log-linear model $[AB]$ $[AC]$ $[BC]$ fits all three two-way margins exactly, so the fitting process must attempt to move the partitions to balance the forces in these margins; hence, the density of cross-hatching in the two-way marginals must be made equal in all three views, as they are in the left column of Figure 8. Badness-of-fit will then be shown by the degree to which the density of points in the *cells* are unequal, as they are in the right column.

Figure 9 shows the three steps in the first iteration (cycle 1) of IPF in the same three-view format. Step 1, shown

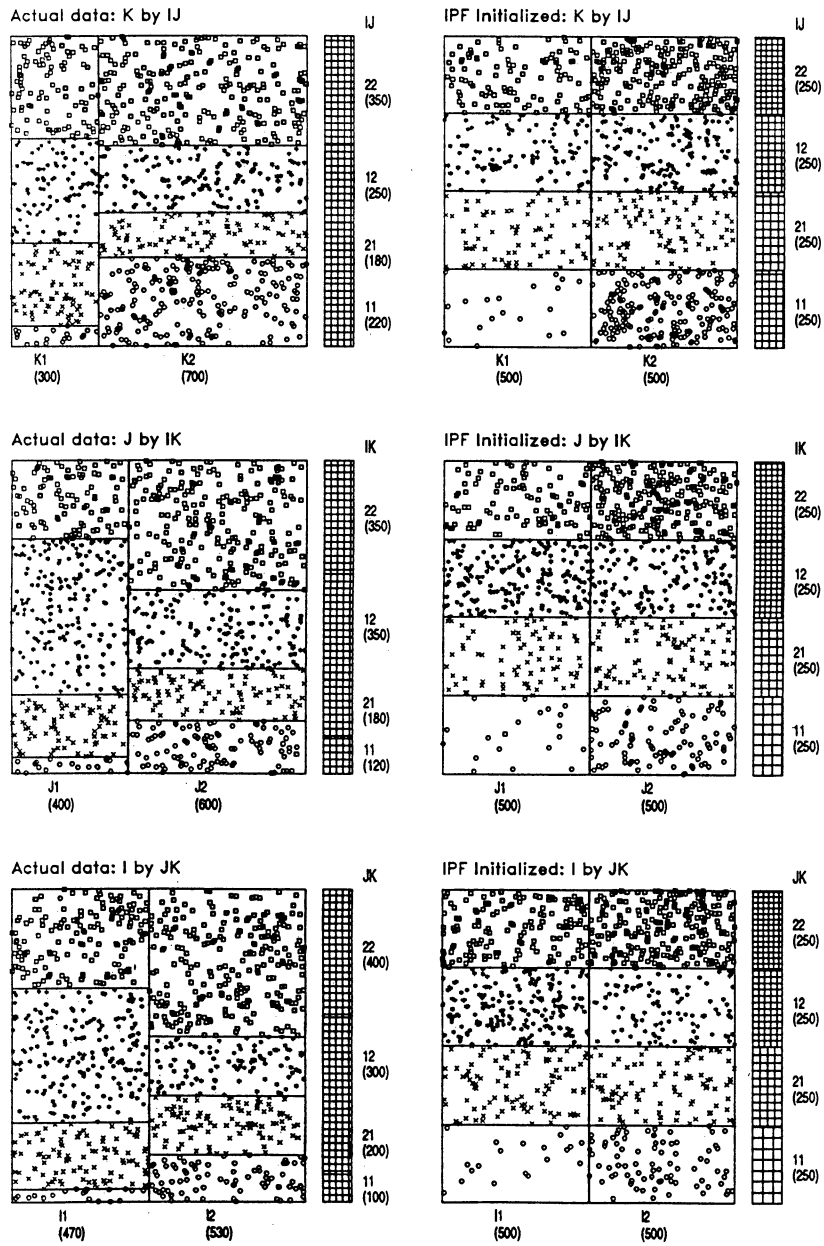

Figure 8. Iterative Proportional Fitting: Data and Initial Estimates. The goal is to make the forces balance in the IJ, IK, and JK margins, as shown by equal cross-hatch density in the margins for all three views of the data.

in the left column, fits the IJ margin. So the forces balance, and cross-hatch density is the same for all IJ margins in the top panel; but the densities are unequal in the two remaining views in this column. Note also that the one-way I and J fitted margins are now equal to those of the data, but the K marginal is not.

In step 2, shown in the middle column, the IK marginals are fitted, which makes the cross-hatch densities equal in the middle view in this column and makes the one-way K fitted marginal match the data. The fitted JK margins are still unequal to those in the data. Finally, the JK margin is fit in the third step (right column) and the cross-hatch densities are now equal in the bottom view. Note that the IJ and IK fitted margins are not preserved, although the differences of their values in step 2 are quite small. These steps are then repeated, reequating the IJ, IK, and JK margins in turn, until the maximum deviation from the observed margins is less than some prescribed amount. In concluding this section, note that the badness-of-fit G^2 (the energy required to enforce the constraints), as judged from the tendency of the cells to vary in point density, decreases dramatically from the initial IPF estimates (right column in Figure 8) to the values in step 3 of the first iteration.

3.4 Numerical Minimization

Methods for minimizing the negative log-likelihood numerically can also be interpreted in terms of our mechanical model (Farebrother 1988). Given an estimate, $\beta^{(t)}$ of the model parameters, where $\widehat{\mathbf{m}}^{(t)} = \exp(\mathbf{X}'\beta^{(t)})$, the method of steepest descent finds an improved estimate $\beta^{(t+1)}$ by moving a small distance λ in the direction of the score vector $\mathbf{f}^{(t)} = \partial \log L/\partial\beta = \mathbf{X}'(\mathbf{n} - \mathbf{m}^{(t)})$ to give $\beta^{(t+1)} = \beta^{(t)} + \lambda\mathbf{f}^{(t)}$. But $\mathbf{f}^{(t)}$ is just the vector of forces in

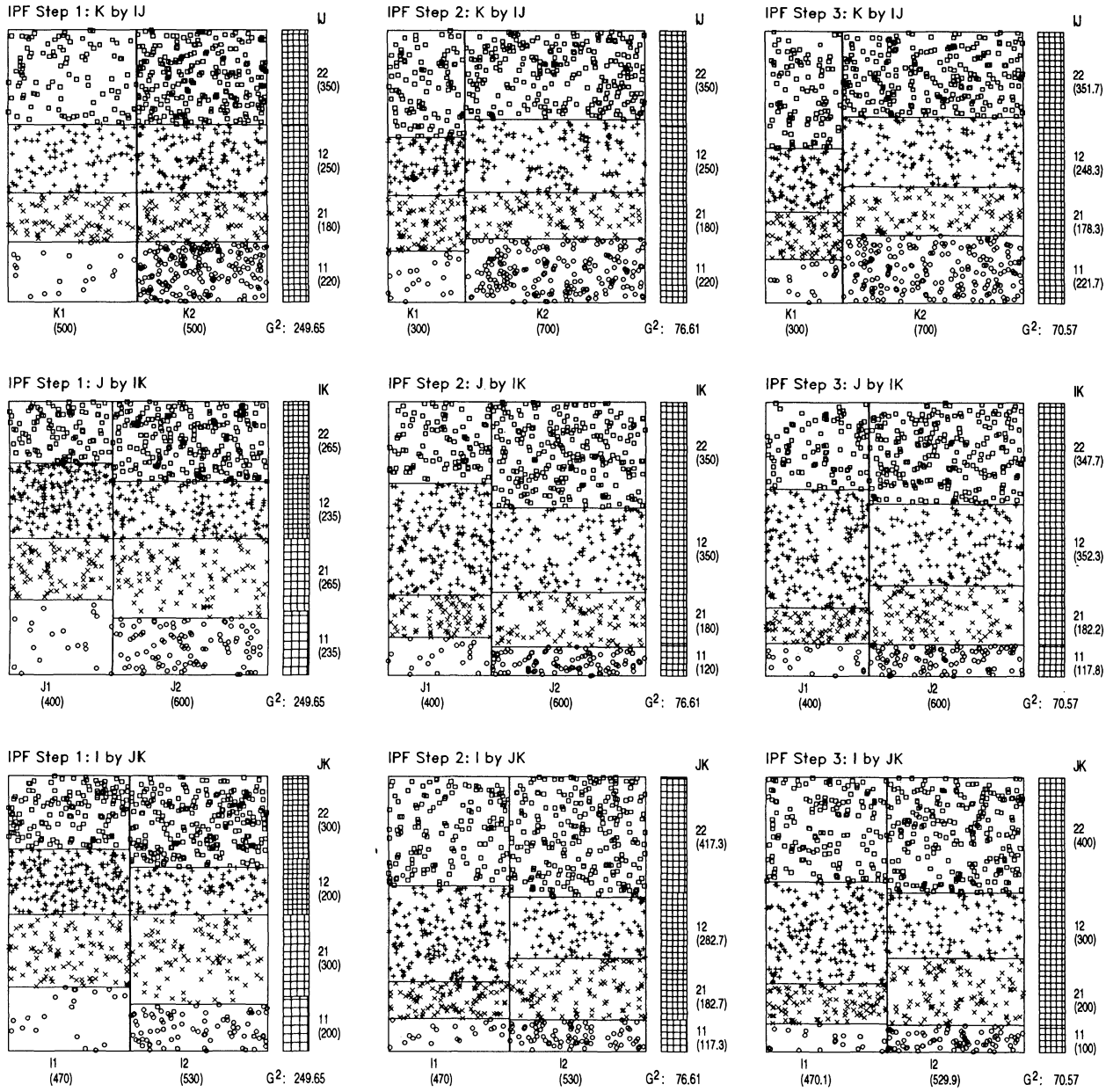

Figure 9. First IPF Iteration. Step 1 fits the IJ margin, making the forces balance in the top-left view; step 2 fits the IK margin, equating forces in the J × IK view; step 3 fits the JK margin in the bottom right view.

the mechanical model attributed to the differences between $\mathbf{n}$ and $\mathbf{m}^{(t)}$ with respect to the parameters. The Newton–Raphson method is similar, but uses an update step $\beta^{(t+1)} = \beta^{(t)} + \mathbf{H}^{(t)}\mathbf{f}^{(t)}$, where $\mathbf{H}^{(t)} = -(\mathbf{X}' \operatorname{diag} \mathbf{m}^{(t)}\mathbf{X})^{-1}$, based on a quadratic approximation to the surface of the energy function (negative log-likelihood) applied to the force vector.

3.5 Detecting Patterns of Association

The mosaic display shows the observed counts in a frequency table by the area of each cell. To see the nature of association, we must imagine what the pressures would be under a null or baseline model as in Figure 6. This is difficult to do unless the observed and expected frequencies can be compared visually.

Friendly's (1991, 1993) enhanced mosaic does this by shading the cells with color and density in relation to the sign and magnitude of the (standardized) residual, $d_{ij} = (n_{ij} - \hat{m}_{ij})/\sqrt{\hat{m}_{ij}}$, from the model (Fig. 10). The sizes of the cells show the data, which corresponds to the saturated model; the shading shows the pattern of residuals, the discrepancies between the data and the model which has been fit. Fitting a nonnull model then moves some of the association from the residuals to the fit, so the process of finding an acceptable model can be thought of as "cleaning the mosaic."

This mixes the metaphor, since area and visual density are used to show different aspects of the data. The combination, however, gives a graphic display that helps the viewer see the data, detect patterns, and suggest hypotheses. Thus, Figure 10 shows that the association between hair color and eye color arises from the high frequencies in the cells for brown eyes and black hair, green eyes and red hair, and blue eyes and blond hair.

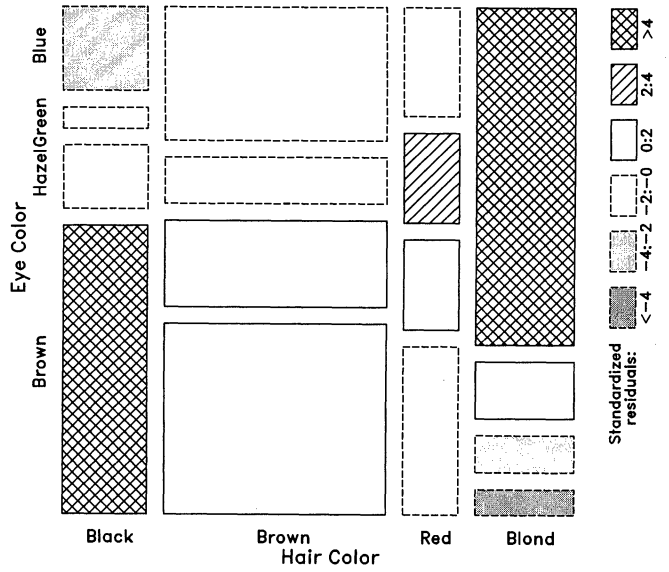

Figure 10. Mosaic Display for Hair Color, Eye Color Data. The area of each tile is proportional to cell frequency as in Figure 5. The style and density of shading, however, is designed to depict the standardized residual from independence, d_{ij}. Cells with $|d_{ij}| < 2$ are empty; cells with $|d_{ij}| \geq 2$ or ≥ 4 are shaded with two levels of density. Positive residuals are outlined with solid lines and filled with cross-hatching; negative residuals are outlined with broken lines. In addition, the eye colors have been permuted to put residuals of like sign in opposite corners in order to better display the pattern of association.

4. CONCLUSION

For some time I have wondered why graphical methods for categorical data are so poorly developed and little used compared with methods for quantitative data. What has made this contrast puzzling is the fact that the statistical methods for categorical data, such as log-linear models and logistic regression, are in most respects discrete analogs of corresponding methods of analysis of variance and regression for quantitative data. One explanation suggested by the conceptual model described here is that categorical data require a different graphic metaphor, and hence a different visual representation from that which has been useful for quantitative data. However, both find common ground in the physical balancing of forces and minimizing potential energy.

A second explanation for this disparity is that graphical methods for quantitative data have been generalized from two variables to three or more variables in several ways. The scatterplot matrix and partial regression residual plots, for example, are natural generalizations of the simple scatterplot designed for multivariate quantitative data. In contrast, other graphical techniques for categorical data tend to be quite specialized. As shown here, the conceptual model for categorical data and the mosaic display generalize readily to multiway tables.

However, it is of some interest to note that most other graphical methods for categorical data [see Friendly (1992)] also depict frequencies by area or visual density. It is hoped that the recognition of an underlying conceptual model for these displays will spur the development and use of graphics for discrete data.

More generally, graphs, like other forms of communication, serve different purposes. The most common goals are summarization and exposure. We also need to think about conceptual models behind the picture. With a powerful conceptual model, a graph can also become a tool for thinking.

[*Received July 1992. Revised September 1994.*]

REFERENCES

Bertin, J. (1983), *Semiology of Graphics* (trans. W. Berg), Madison, WI: University of Wisconsin Press.

Farebrother, R. W. (1987), "Mechanical Representations of the L1 and L2 Estimation Problems," in *Statistical Data Analysis Based on the L1 Norm and Related Methods*, ed. Y. Dodge, Amsterdam: North-Holland, pp. 455–464.

——— (1988), "On an Analogy Between Classical Mechanics and Maximum Likelihood Estimation," *Osterrische Zeitschrift fur Statistik und Informatik*, 18, 303–305.

Friendly, M. (1991), "Mosaic Displays for Multi-Way Contingency Tables," York Univ., Dept. of Psychology Reports, No. 195.

——— (1992), "Graphical Methods for Categorical Data," in *Proceedings of the SAS User's Group International Conference*, 17, pp. 1367–1373.

——— (1993), "Mosaic Displays for Multi-Way Contingency Tables," *Journal of the American Statistical Association*, 89, 190–200.

Good, I. J. (1963), "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables," *Annals of Mathematical Statistics*, 34, 911–933.

Sall, J. (1991a), "The Conceptual Model Behind the Picture," *ASA Statistical Computing and Statistical Graphics Newsletter*, 2, 5–8.

——— (1991b), "The Conceptual Model for Categorical Responses," *ASA Statistical Computing and Statistical Graphics Newsletter*, 3, 33–36.

Snee, R. D. (1974), "Graphical Display of Two-Way Contingency Tables," *The American Statistician*, 28, 9–12.