

ESTIMATING FLEXIBLE DISTRIBUTIONS OF IDEAL-POINTS WITH EXTERNAL ANALYSIS OF PREFERENCES

WAGNER A. KAMAKURA

VANDERBILT UNIVERSITY

Ideal-points are widely used to model choices when preferences are single-peaked. Ideal-point choice models have been typically estimated at the individual-level, or have been based on the assumption that ideal-points are normally distributed over the population of choice makers. We propose two probabilistic ideal-point choice models for the external analysis of preferences that allow for more flexible multimodal distributions of ideal-points, thus acknowledging the existence of subpopulations with distinct preferences. The first model extends the ideal-point probit model for heterogeneous preferences to accommodate a mixture of multivariate normal distributions of ideal-points. The second model assumes that ideal-points are uniformly distributed within finite ranges of the attribute space, leading to a more simplistic formulation and a more flexible distribution. The two models are applied to simulated and actual choice data, and compared to the ideal-point probit model.

Key words: choice models, multinomial probit.

Introduction

Since Coombs (1964) unfolding model, ideal-points have been widely used to represent individual preferences along an attribute space. Coombs' objective was to "unfold" preference rankings from a sample of subjects to obtain a joint space of stimuli and subjects (represented by their ideal stimulus) so that the distances from ideal-points to stimuli closely relate to the observed preference rankings. This original unidimensional model was later extended to the multidimensional case by Bennet and Hays (1960) and Hays and Bennet (1961).

Although the objective in unfolding models is to obtain the location of both the subject ideal-points and stimuli in a joint space, one often finds situations in which the researcher's interest focuses on the location of ideal-points in a pre-existing space defined by predetermined attributes. The location of consumers ideal-points on certain physical attributes, for example, is useful information for the development of new products that satisfy the needs of a group of consumers. This process of mapping preferences for a known set of stimuli in a predefined space was named the external analysis of preferences by Carroll (1972), in contrast to the internal analysis of preferences by unfolding models.

The external analysis of preferences with ideal-point models is especially well-suited when preferences are being directly related to psychophysical attributes, such as sweetness, sourness, and temperature, where preferences are expected to be single-peaked (Coombs & Avrunim, 1977). Srinivasan and Shocker's (1973) linear programming procedure for the external analysis of preferences includes an ideal-point preference function to accommodate these applications. Although their LINMAP procedure estimates a deterministic preference function based on observed preference rankings,

This research was funded by the Dean's Fund for Faculty Research of the Owen Graduate School of Management, Vanderbilt University.

Request for reprints should be sent to Wagner A. Kamakura, Vanderbilt University, 401 21st Avenue South, Nashville, Tennessee 37203.

Shocker and Srinivasan (1974) grant that distances between ideal-point and stimuli should be associated with probabilities of purchase rather than deterministic events.

Cooper and Nakanishi's (1983) ideal-point logit model for the external analysis of preferences acknowledges that stated preferences or observed choices may, at best, possess strong stochastic transitivity (Tversky & Russo, 1969). Their model, however, prescribes a single ideal-point for a group of subjects, thus assuming perfect homogeneity of preferences within the group, which results in the property of independence from irrelevant alternatives (IIA), unless the model is applied to various predetermined homogeneous groups of subjects.

Since ideal-point models are especially applicable to situations involving individual tastes for psychophysical attributes, one may expect intra- and inter-individual differences in preferences, even within fairly homogeneous groups. Rather than specifying a fixed ideal-point, Kamakura and Srivastava (1986) propose a model that estimates the aggregate distribution of ideal-points within a group, based on observed choices. They also demonstrate that this ideal-point probit model is not constrained by the IIA property. Choice share predictions for an alternative depend not only on its proximity to the mean ideal-point, but also on its proximity to other alternatives so that similar alternatives (closely placed on the attribute space) tend to be substitutable.

Similar assumptions are made by De Soete, Carroll, and DeSarbo (1986) in their probabilistic unfolding model. They also demonstrate that with these assumptions, the choice model does not imply strong stochastic transitivity as Cooper and Nakanishi's (1983), but satisfies the more desirable moderate stochastic transitivity (Halff, 1976). Despite this desirable property, these latter models make the assumption that ideal-points are normally distributed (i.e., clustered around a single group mean). This assumption ignores the possibility of distinct preference groups that would lead to a multimodal distribution.

In this paper we propose two ideal-point probabilistic choice models that allow for more flexible distributions of ideal-points. The first model extends the Kamakura and Srivastava (1986) probit model to multimodal distributions of ideal-points; the second is based on several simplifying assumptions that lead to a multidimensional histogram of ideal-points. The next section outlines the original probit model with a multivariate normal distribution of ideal-points. Although a detailed description of this probabilistic choice model is found elsewhere (Kamakura & Srivastava), it is outlined here to provide the background for the discussion of the multimodal models. Following that, this model is extended to a multimodal distribution combining several normal random variables. Next, we derive the histogram model, and present two empirical illustrations based on simulated and actual data.

The Ideal-Point Probit Model

Let us first define:

- C: a choice set consisting of alternatives $j = 1, 2, \dots, J$;
- X: a K by J matrix with the coordinates of the J alternatives in a K -dimensional attribute space;
- z: a K by 1 vector defining the location of the individual's ideal-point in the same K -dimensional attribute space; and
- ω : a K by K diagonal matrix containing the relative importance weights of each attribute.

The choice process assumed is the same as in Kamakura and Srivastava (1986) and DeSoete, Carroll, and Desarbo (1986). At each choice occasion, the choice maker is

assumed to draw a random ideal-point \mathbf{z} from a given multivariate normal distribution (with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$), and assign propensities to the choice alternatives according to how close they are to the drawn ideal-point. The choice probability for alternative $j \in C$ is given by

$$P_j = \Pr [D_j + \varepsilon_j < D_m + \varepsilon_m; \forall m \in C; m \neq j], \quad (1)$$

where D_j is the weighted squared-Euclidean distance between alternative j and the random ideal-point,

$$D_j = (\mathbf{X}_j - \mathbf{z})' \boldsymbol{\omega} (\mathbf{X}_j - \mathbf{z}), \quad (2)$$

where \mathbf{X}_j is the K by 1 vector of coordinates for alternative j , and ε_j are iid normal random disturbances with mean zero and standard deviation σ that account for other sources of randomness in choice behavior (Kamakura & Srivastava, 1986, pp. 202–3).

One can view the random variable D_j as a latent “disutility” assigned to the choice alternative j at a particular choice occasion. Setting the weight matrix $\boldsymbol{\omega}$ equal to an identity matrix will result into the unweighted ideal-point model (Carroll, 1972, 1980; Coombs, 1964; De Soete, Carroll, & DeSarbo, 1986).

Departures from the ideal-point along some attributes might be more critical to the choice-makers than along others, justifying the inclusion of the diagonal matrix of importance weights $\boldsymbol{\omega}$. A diagonal matrix $\boldsymbol{\omega}$ having positive entries on the main diagonal will lead to iso-preference surfaces with an ellipsoidal shape. A negative weight ω_{kk} will lead to an “anti-ideal-point” (Carroll, 1972); choice-makers would prefer alternatives away from the ideal-point, along attribute k . A few algebraic manipulations of (1) and (2) yields

$$P_j = \Pr [2(\mathbf{X}_m - \mathbf{X}_j)' \boldsymbol{\omega} \mathbf{z} + (\varepsilon_j - \varepsilon_m) < \mathbf{X}_m' \boldsymbol{\omega} \mathbf{X}_m - \mathbf{X}_j' \boldsymbol{\omega} \mathbf{X}_j; \forall m \in C; m \neq j]. \quad (3)$$

Because \mathbf{z} is distributed as a multivariate normal random variable with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and the ε_j 's are assumed independent normal random variables with mean zero and standard deviation σ , the first term within brackets,

$$\eta_m = 2(\mathbf{X}_m - \mathbf{X}_j)' \boldsymbol{\omega} \mathbf{z} + (\varepsilon_j - \varepsilon_m), \quad (4)$$

is also normally distributed with mean

$$E\{\eta_m\} = 2(\mathbf{X}_m - \mathbf{X}_j)' \boldsymbol{\omega} \boldsymbol{\mu}, \quad (5)$$

and covariance matrix $\boldsymbol{\Omega}$, with main-diagonal entries of

$$\Omega_{mm} = 4(\mathbf{X}_m - \mathbf{X}_j)' \boldsymbol{\omega} \boldsymbol{\Sigma} \boldsymbol{\omega} (\mathbf{X}_m - \mathbf{X}_j) + 2\sigma^2, \quad (6)$$

and off-diagonal entries of

$$\Omega_{mm'} = 4(\mathbf{X}_m - \mathbf{X}_j)' \boldsymbol{\omega} \boldsymbol{\Sigma} \boldsymbol{\omega} (\mathbf{X}_m - \mathbf{X}_j). \quad (7)$$

The choice probabilities are thus obtained from a covariance probit model,

$$P_j = \Phi\{\mathbf{X}_m' \boldsymbol{\omega} \mathbf{X}_m - \mathbf{X}_j' \boldsymbol{\omega} \mathbf{X}_j; \forall m \in C; m \neq j\}, \quad (8)$$

where $\Phi\{\cdot\}$ represents the cumulative multivariate normal distribution with the means in (5) and covariance matrix $\boldsymbol{\Omega}$. Notice that multiplying the importance weights in $\boldsymbol{\omega}$ by any positive constant has no impact on the probabilities in (8). To eliminate this indeterminacy, the importance weights are constrained to a sum-of-squares of 1. The estimation of this probit model based on binary choice data is discussed in detail in De Soete, Carroll, and DeSarbo (1986) and based on polytomous choices in Kamakura and

Srivastava (1986). The later authors also demonstrate how this multinomial probit model can be estimated based on preference rankings.

As mentioned before, the main limitation of this model is the restrictive assumption that ideal-points are normally distributed over the attribute space, which ignores the possibility of local concentrations of ideal-points representing fairly homogeneous groups of choice makers. In the following sections, we propose an extension of this model for multimodal distributions, and a distinct approach that allows for more flexible distributions of ideal-points.

An Extension For Multimodal Distributions of Ideal-Points

In this extension, we make the assumption that at each observed choice occasion, ideal-points might be drawn from N distinct multivariate normal distributions with means $\boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{\Sigma}_i$. The choice probabilities $P_{j|i}$, conditional on the particular distribution of ideal points, are obtained from (3) to (7), given the parameters of the multivariate normal distribution of ideal-points:

$$P_{j|i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i, \mathbf{X}) = \Phi\{\mathbf{X}'_m \boldsymbol{\omega}_i \mathbf{X}_m - \mathbf{X}'_j \boldsymbol{\omega}_i \mathbf{X}_j; \forall m \in C; m \neq j\}. \quad (9)$$

Conditional choice probabilities, given a particular subpopulation of ideal-points i , are thus obtained from the wandering ideal-point model and satisfies its properties. The unconditional choice probabilities are determined by a mixture of these N probit models according to a multinomial process that determines the relative frequency of the N subpopulations of ideal-point distributions (f_i):

$$P_j = \sum_{i=1}^N \{P_{j|i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i, \mathbf{X}) f_i\}; \sum_{i=1}^N f_i = 1.$$

We use a logit formulation to model the relative frequency of the subpopulations, $f_i = \exp(\theta_i) / \sum_k \exp(\theta_k)$, which leads to a mixture of a multinomial logit model with probit conditional probabilities,

$$P_j = \sum_{i=1}^N \left\{ \frac{P_{j|i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i, \mathbf{X}) \exp(\theta_i)}{\sum_{k=1}^N \exp(\theta_k)} \right\}. \quad (10)$$

This model requires the estimation of $(3K - 1) + K(K - 1)/2$ probit parameters for each subpopulation i (i.e., K elements of the mean vector $\boldsymbol{\mu}_i$, $K(K - 1)/2 + K$ covariance terms in $\boldsymbol{\Sigma}_i$ and the $(K - 1)$ importance weights in $\boldsymbol{\omega}_i$), in addition to the $(N - 1)$ logit parameters θ_i ($\theta_N = 0$).

Estimating the Multimodal Ideal-Point Model

Define a Boolean variable y_{jt} with value 1 if alternative j is chosen at occasion t , ($t = 1, 2, \dots, T$) or 0 otherwise. The log likelihood (LL) for a sample of observed choices on T choice occasions will then be

$$LL = \sum_{t=1}^T \{y_{jt} \ln [\boldsymbol{\Sigma}_i \exp(\theta_i) P_{j|i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_i, \mathbf{X}_t)] - y_{jt} \ln [\boldsymbol{\Sigma}_i \exp(\theta_i)]\}. \quad (11)$$

The first and second derivatives of the log-likelihood function relative to the frequency and probit parameters can be easily computed. Maximum likelihood estimates of the probit parameters ($\mu_i, \Sigma_i, \omega_i$) and the logit parameters θ_i , based on observed choices, can thus be obtained via a Newton-Raphson search. Another alternative estimation procedure for mixture models such as the one proposed here is the EM algorithm. Details on the estimation of mixture models using the EM algorithm can be found in McLachlan and Basford (1988).

The model described above, when applied to the observed choices from a random sample of the population of consumers, will locate the distribution of ideal-points for each of N subpopulations (probit parameters) and estimate the relative size of these subpopulations (θ_i). The number of subpopulations (N) may be determined by estimating the unimodal, bimodal, etcetera models sequentially and then using Akaike's AIC criterion, or a Bayesian cross-validated likelihood test (Fornell & Rust, 1989) to select the most adequate value for N . For example, let L_n ($n = 1, 2, \dots, N$) be the log likelihood attained with n subpopulations and M parameter estimates. Akaike's AIC criterion for the n -groups model is given by

$$AIC_n = -2(L_n - M). \quad (12)$$

The AIC criterion, however, relies on the same regularity conditions of the likelihood-ratio test (McLachlan & Basford, 1988) and should thus be used only as a qualitative indicator.

Assuming there is no prior theoretical support for a particular number of subpopulations n , the posterior probability that the n -groups model is the correct one is given by (Fornell & Rust, 1989)

$$BCVL_n = \left[\sum_{l=1}^N \exp(L_l - L_n) \right]^{-1}. \quad (13)$$

The Histogram Model

We start with the same basic assumptions: at any choice occasion, a random ideal-point is drawn from a distribution, and the alternative closest to the ideal is chosen by the decision maker. However, rather than assuming a multivariate normal distribution of the ideal-points, or a mixture of normal distributions, we assume that ideal-points are uniformly distributed within prespecified ranges of attributes. In other words, we first partition the attribute space into cells and assume that these cells are reasonably small so that within these narrow ranges of the attributes, the distribution of ideal-points can be approximated with a uniform distribution. Our objective is thus to estimate the frequency distribution of ideal-points ($f_i, i = 1, 2, \dots, N$) across all N rectangular cells of the prespecified grid.

Conditional Probabilities of Choice Within a Rectangular Cell

To simplify our derivations, let us first consider choices between only two alternatives (the extension to polytymous choices will be discussed next). Since we assume that the alternative closest to the drawn ideal-point is chosen, it follows from (2) that if alternative A is chosen over B ,

$$g(\mathbf{X}, \boldsymbol{\omega}, \mathbf{z}) = 2(\mathbf{X}_B - \mathbf{X}_A)' \boldsymbol{\omega} \mathbf{z} - \mathbf{X}_B' \boldsymbol{\omega} \mathbf{X}_B + \mathbf{X}_A' \boldsymbol{\omega} \mathbf{X}_A < 0, \quad (14)$$

and if B is chosen over A , $g(\mathbf{X}, \boldsymbol{\omega}, \mathbf{z}) > 0$. Therefore, one could derive the conditional choice probability for alternative A within a cell by determining the cumulative distribution function of the weighted average of uniformly distributed random variables $\nu = 2(\mathbf{X}_B - \mathbf{X}_A)' \boldsymbol{\omega} \mathbf{z}$ and evaluating this cumulative distribution function at the point $\mathbf{X}'_B \boldsymbol{\omega} \mathbf{X}_B - \mathbf{X}'_A \boldsymbol{\omega} \mathbf{X}_A$. However, since we assume that the ideal-points are uniformly distributed within cells, one can easily compute choice probabilities using simple geometry, as shown below.

Function $g(\cdot)$ partitions the attribute space in two regions so that any random ideal-point location \mathbf{z} that yields $g(\mathbf{X}, \boldsymbol{\omega}, \mathbf{z}) < 0$ would lead to a choice of alternative A , or B if it yields $g(\mathbf{X}, \boldsymbol{\omega}, \mathbf{z}) > 0$. Now, let \mathbf{x}_i^L and \mathbf{x}_i^H represent the lower and upper boundaries of cell i in the attribute space. Since ideal-points are uniformly distributed in this region and the alternative closest to the random ideal-point is chosen, all choices of alternative A from this cell must come from ideal-points located in the polyhedron Δ_A defined by

$$\mathbf{x}_i^L < \mathbf{z} < \mathbf{x}_i^H,$$

$$g(\mathbf{X}, \boldsymbol{\omega}, \mathbf{z}) < 0.$$

Similarly, all choices of alternative B from cell i must come from ideal-points located in the polyhedron Δ_B defined by

$$\mathbf{x}_i^L < \mathbf{z} < \mathbf{x}_i^H,$$

$$g(\mathbf{X}, \boldsymbol{\omega}, \mathbf{z}) > 0.$$

One can easily see that the polyhedrons Δ_A and Δ_B remain unchanged whenever the matrix of importance weights $\boldsymbol{\omega}$ is multiplied by a positive constant. This indeterminacy is eliminated by constraining the diagonal elements of $\boldsymbol{\omega}$ to sum to one (if they are all assumed to be nonnegative), or by constraining their sum-of-squares to one (if anti-ideal points are allowed).

The conditional choice probability for alternative A over alternative B within cell i , $P_i[A, B]$ is the ratio between the volume of Δ_A and the sum of the volumes for Δ_A and Δ_B . Extending this computation to polytomous choices is straightforward. In the trichotomous case, one would need three isopreference functions similar to $g(\cdot)$, to define the regions of dominance for each choice alternative, within the cell.

Estimation of the Cell Frequencies

Given the location of the choice alternatives (\mathbf{X}), importance weights ($\boldsymbol{\omega}$), and a prespecified cell structure, the conditional probabilities of choice within each cell i ($P_{j|i}$), can be computed as described above. The unconditional choice probability for alternative j is then given by

$$P_j = \sum_i f_i P_{j|i}, \quad (15)$$

where $f_i = \exp(\theta_i) / \sum_k \exp(\theta_k)$ represents the frequency of ideal-points in cell i .

Maximum likelihood estimates of θ_i based on observed choices over choice occasions t ($t = 1, 2, \dots, T$) are obtained by finding the minimum of the nonlinear function:

$$LL = \sum_t y_{jt} \ln [\sum_i \exp(\theta_i) P_{j|i}] - \sum_t y_{jt} \ln \left[\sum_i \exp(\theta_i) \right], \quad (16)$$

where y_{jt} is the Boolean variable indicating whether alternative j was chosen at occasion t . Since the conditional within-cell probabilities do not depend on the frequency parameters θ_j , the first and second derivatives of the log-likelihood function relative to the frequency parameters are fairly simple. Therefore, ML estimates of the cell frequencies in the unweighted ideal-point model ($\omega = \mathbf{I}_K$) can be obtained through a Newton-Raphson search.

The analytical form of the derivatives for the estimation of the importance weights ω is much more complex, since it involves the conditional choice probabilities $P_{j|i}$, which were computed geometrically. Given this complexity, and considering that only a few ($K - 1$) importance weights are estimated, the gradient of the log-likelihood function may be computed via numerical approximations.

Estimation of Cell Frequencies from Aggregate Choice Data

Estimating the histogram of ideal-points based on aggregate choice shares (rather than discrete choice observations) is much simpler and faster. Let, $M = J - 1$,

- \mathbf{P}_t : ($N \times M$) matrix of conditional choice probabilities for the alternatives (except for one) considered at occasion t ;
- \mathbf{S}_t : ($M \times 1$) vector of observed choice shares at occasion t ;
- \mathbf{f} : ($N \times 1$) vector of frequencies for the N cells, to be estimated.

Least-squares estimates of the cell frequencies \mathbf{f} will be given by

$$\min_{\mathbf{f}} \left(\frac{1}{T} \right) \Sigma_t \text{tr} \{ (\mathbf{S}_t - \mathbf{P}_t' \mathbf{f})' (\mathbf{S}_t - \mathbf{P}_t' \mathbf{f}) \}, \tag{17}$$

$$\min_{\mathbf{f}} - \left(\frac{2}{T} \right) \Sigma_t \text{tr} \{ \mathbf{S}_t' \mathbf{P}_t' \mathbf{f} \} + \left(\frac{1}{T} \right) \Sigma_t \mathbf{f}' \mathbf{P}_t \mathbf{P}_t' \mathbf{f} + \left(\frac{1}{T} \right) \Sigma_t \text{tr} \{ \mathbf{S}_t' \mathbf{S}_t \}, \tag{18}$$

or, after dividing by 2 and subtracting $(\frac{1}{T}) \Sigma_t \text{tr} \{ \mathbf{S}_t' \mathbf{S}_t \}$,

$$\min_{\mathbf{f}} - \left(\frac{1}{T} \right) \left[\Sigma_t \text{tr} \{ \mathbf{S}_t' \mathbf{P}_t' \mathbf{f} \} + \left(\frac{1}{2} \right) \mathbf{f}' (\Sigma_t \mathbf{P}_t \mathbf{P}_t') \mathbf{f} \right], \tag{19}$$

subject to $\mathbf{1}' \mathbf{f} = 1$; $\mathbf{f} \geq \mathbf{0}$.

This quadratic programming problem will yield the least-square estimates of the cell frequencies (\mathbf{f}) for the unweighted ideal-point model.

The sampling properties of constrained least-squares estimators such as the one presented above are discussed in Judge, Griffiths, Hill, and Lee (1980). With regard to equality constraints (such as $\mathbf{1}' \mathbf{f} = 1$), these authors conclude that "It is apparent . . . that if the outside information [about the constraints] is correct, then using it in conjunction with the sample information will lead to an unbiased estimator that has a precision matrix superior to the unrestricted least-squares estimator" (p. 58). As for inequality constraints (such as $\mathbf{f} \geq \mathbf{0}$), these authors conclude that if the researcher knows that a parameter is nonnegative (as in our case), the inequality-restricted estimator is uniformly superior to the unrestricted estimator (Judge, et al. 1980, p. 93). However, when applying this simple least-squares estimation procedure, one must be aware of the fairly large number of parameters required by the model (i.e., the vector of frequencies \mathbf{f}), which will depend on the particular cell structure specified a priori. Because of the large number of parameters to be estimated, this aggregate estimation procedure is limited to choice experiments involving a large number of choice decisions

among a variety of choice sets, such as the paired comparisons used in our empirical tests (to be discussed next).

The estimation of the weighted ideal-point model will require a nonlinear least-squares procedure, because the conditional choice probability $P_{j|i}$ within cell i is nonlinear on the importance weights ω . The algorithm developed for estimating the weighted model uses a gradient search for the importance weights, using the sum-of-squared residuals of the quadratic programming solution in (19) as the objective function to be minimized.

Empirical Illustrations

The two extensions of the random ideal-point model are illustrated here with the use of simulated data and data collected through a laboratory experiment. The simulations were conducted to verify the ability of the estimation algorithms to uncover the true distribution of ideal-points. The results from the laboratory experiment will indicate whether the models are able to explain actual preferences.

Simulated Data

A trimodal distribution of ideal-points in a two-attribute space was first created, by combining three bivariate normal distributions with means $(\mu_{i1}, \mu_{i2}; i = A, B, C)$, standard deviations $(\sigma_{i1}, \sigma_{i2}; i = A, B, C)$ correlations $r_{i12} = 0$, and specifying that 35% of all ideal-points are generated from the bivariate normal A , 35% from B , and 30% from C . Samples of 1000 ideal-points from the three subpopulations were generated by varying the standard deviations σ_{i1} and σ_{i2} .

An experimental design consisting of 16 stimuli that combined 4 levels (1.0, 2.0, 3.0, and 4.0) of the two attributes was then defined. Two conditions were used to generate choice shares for all 120 pairs of stimuli, assuming equal attribute importances. In the first condition (simulations I, II, III, and IV), choices were simulated by assigning each of the 1000 ideal-points to the closest stimulus of the pair, following the underlying assumptions of the histogram model. In the second condition (simulations V, VI, VII, and VIII), a normally distributed disturbance was added to the "disutility" in (2), representing an additional randomness in choice behavior at each choice occasion, thus violating the underlying assumptions of the histogram model.

The simulated choices within each paired comparison, along with attribute levels for the 16 stimuli were used as inputs for estimating the models—aggregate pairwise shares were used in these simulations for convenience (since it would not change the likelihood function); both the histogram and probit models may be estimated based on polytomous choices. Additional choice shares were computed for all pairwise combinations of 5 other stimuli, used for hold-out prediction tests.

The actual and estimated distributions for one of the simulations (I) are displayed in Figure 1. One can see that the bimodal probit attempted to capture the heterogeneity in preferences by fitting one mode to group A and the other between groups B and C . Overall, the estimated histograms with 25 prespecified cells portrayed the actual distribution of ideal-points better than the unimodal and bimodal probit.

As one would have expected, the trimodal probit model retrieved the true distribution of ideal-points fairly well, especially for the first condition (simulations I, II, III, and IV). Even with the inclusion of a random disturbance to the data (simulations V, VI, VII, and VIII), the recovery of the mean ideal-points within each subgroup was reasonably accurate, although substantial bias was found in the estimates of the standard deviations under this second condition due to the violation of the underlying assumptions of the model.

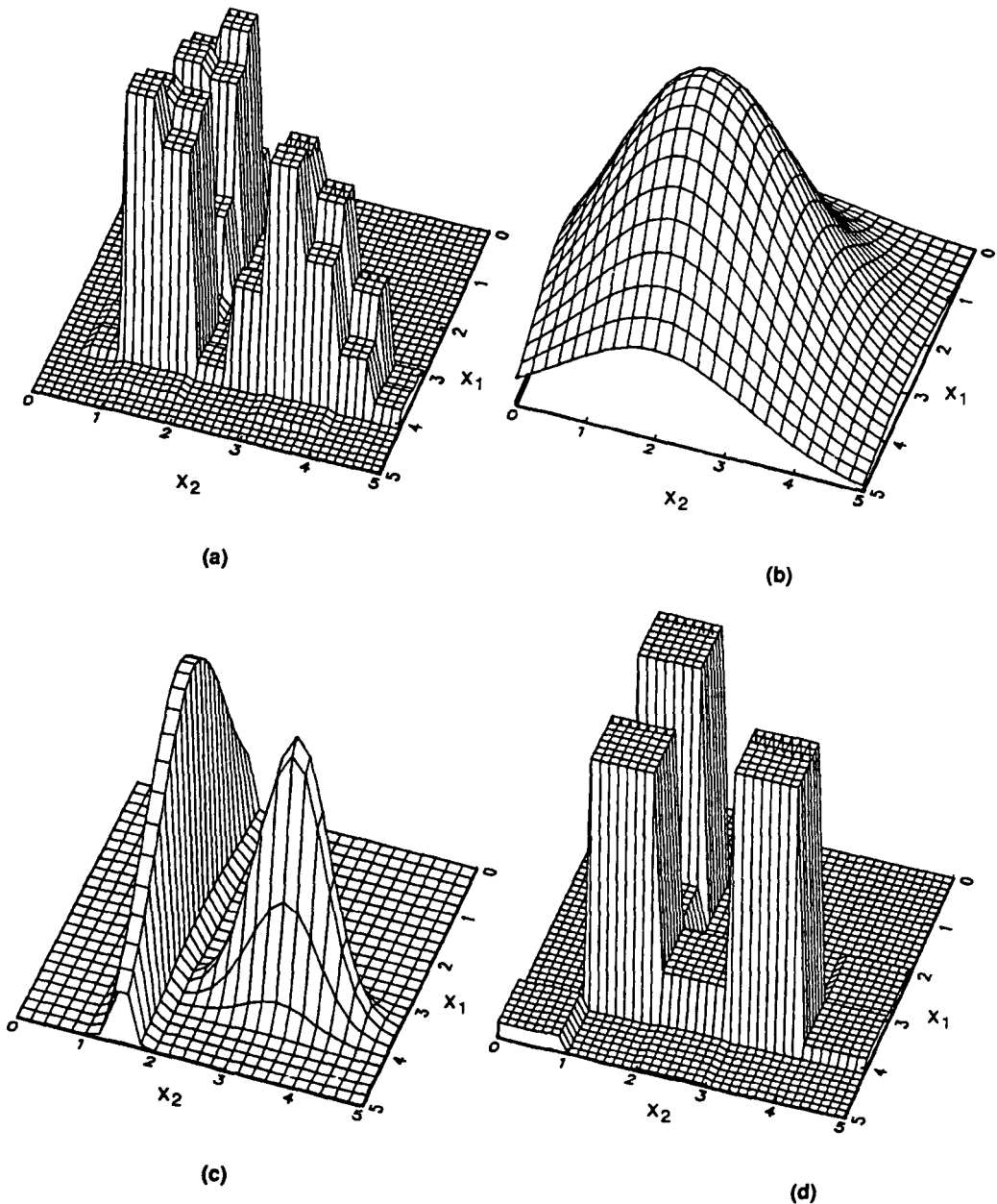


FIGURE 1

Estimates for a trimodal distribution of ideal-points for a representative simulation: (a) actual frequencies; (b) unimodal probit model; (c) bimodal probit model; (d) histogram model.

A comparison of the Akaike information criterion (AIC) obtained with the trimodal, bimodal, and unimodal probit models clearly indicated that trimodal distribution as the most adequate among these formulations of the probit model. Because of the magnitude of the log likelihoods, computation of the BCVL posterior probabilities was unnecessary, since they would accept the trimodal probit model with near certainty.

The bimodal probit model fitted the data reasonably well (R^2 ranging from 0.960 to 0.991 on the aggregate pairwise shares), even though a trimodal distribution was used to generate the choice data. Nevertheless, its performance in terms of fit and predictive

fit was in general inferior to the histogram model. The superior goodness-of-fit of the histogram comes as no surprise, because of the larger number of parameters being estimated (25 cell frequencies, constrained by $\sum_k f_k = 1$ and $f_k \geq 0$) from 120 observations. The superiority in predictive fit would become even more evident, if the distributions of ideal-points were located farther apart. For example, when fairly homogeneous and distinct distributions were used to generate the data (e.g., simulations I, IV, V, and VIII), the superiority of the histogram model over the bimodal probit in terms of predictive fit was stronger (R^2 of 0.937, 0.920, 0.932, and 0.894, for the histogram model, compared to 0.918, 0.771, 0.891, and 0.839, respectively, for the bimodal probit model).

Actual Preference Data

These data were obtained from a replication of an experiment by Batsell (1980). Sixteen prototypes of ordinary notebook pages were constructed by combining the page size (66.5, 80, 93.5, and 107.9 square inches) and the spacing between lines (0.188, 0.25, 0.313, and 0.375 inches) in a full factorial design—a similar data set was previously used by Kamakura and Srivastava (1986) to illustrate their ideal point probit model.

Aggregate choice shares within all pairs of the 16 stimuli were obtained by transforming the preference rankings of 98 subjects into paired comparisons (Chapman & Staelin, 1982). These aggregate shares within 120 pairs, along with the two attributes for each stimuli were used to estimate the unimodal and bimodal probit, and the histogram model. For the sake of parsimony, a common set of importance weights was estimated for both subpopulations ($\omega_A = \omega_B$).

The histogram model was estimated in two stages. In the first stage a coarse grid of 48 cells was specified covering the relevant range of the attribute space (see Figure 2a). Then, a finer grid was specified covering the attribute range with positive frequencies in the first stage. This second stage lead to the estimated distribution in Figure 2b. Based on this histogram, one concludes that ideal-points appear to be concentrated in the range from 65 to 120 sq.in. of area, and from 0.25 to 0.35 inches of line spacing.

Estimation of the unimodal and bimodal probit models lead to the distributions displayed in Figures 2c and 2d, respectively. The unimodal distribution (Figure 2c) is in reasonable agreement with the estimated histogram on Figure 2b, with a mode located at 90.9 sq.in. of size and 0.30 in. of line spacing. The agreement between the histogram estimates and the bimodal distribution is even better. The bimodal distribution in Figure 2d shows the modes at (84.6, 0.31) and (97.5, 0.28), with a total distribution that resembles the histogram quite closely.

Based on the AIC criterion (12,504 for the unimodal and 12,506 for the bimodal model), one concludes that the distribution of ideal-points is unimodal. The BCVL posterior probabilities for the unimodal and bimodal probit models are 0.731 and 0.269, respectively, thus confirming the unimodality of the distribution of ideal-points.

In retrospect, there is a reasonable explanation for this unimodal distribution. One of the stimuli used in the experiment (having size = 93.5 sq.in. and a spacing of 0.313 in.) was quite similar to the regular notebook used by college students, while the other stimuli maintained the same proportions of length and width at different sizes. The respondents familiarity with this combination of size and line spacing may have influenced their answers, leading to a mean ideal-point close to the familiar combination of size and spacing. It is possible that if the length of the page (rather than the area) was varied, covering both letter and legal formats, one would find two preference groups, and consequently, a bimodal distribution of ideal-points.

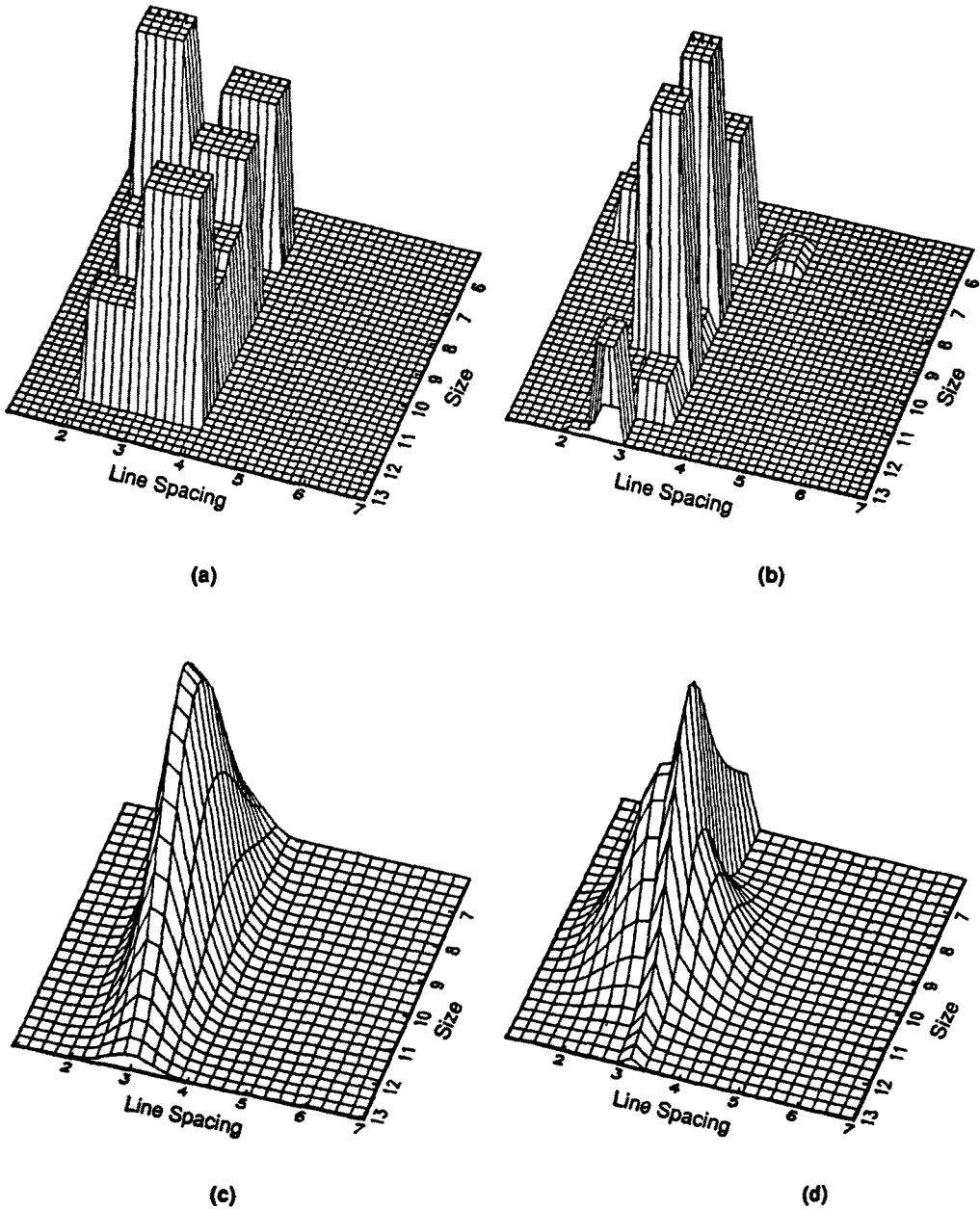


FIGURE 2

Estimates for the notebook experiment: (a) coarse grid histogram; (b) fine grid histogram; (c) unimodal probit model; (d) bimodal probit model.

Discussion

In this paper we have relaxed one assumption of the ideal-point probit model that may be unrealistic in situations where ideal-points form homogeneous clusters in the attribute space. One of the models proposed here is a direct extension of the probit model, with a mixture of multivariate normal distributions. As such, the model is affected by the well-known implementation problems of polytomous probit models, namely, the need for numerical approximations in computing choice probabilities for

more than two-choice alternatives. Since our applications involved pairwise shares, they were not affected by these implementation problems. On the other hand, the histogram model is not affected by such problems. Choice probabilities can be exactly computed, given the underlying assumption of a uniform distribution of ideal-points within a finite cell. Nevertheless, the histogram model requires an ad-hoc definition of a cell structure that justifies the uniform assumption within cells. The heuristic we used in determining this cell structure was to first define a coarse cell structure covering the relevant range of the attributes, estimate the cell frequencies, and define a finer structure covering the cells with positive frequencies. Obviously, the finer the cells' structure specified a priori, the larger the number of cell frequencies to be estimated.

Both the multimodal probit and the histogram models estimate flexible distributions of ideal-points directly from observed choice or preference data. Another alternative for the estimation of ideal-point densities has been suggested by Donthu (1986), who first uses well-known unfolding methods to obtain subject-level estimates of ideal-points, and then overlays a kernel density function on these estimates. In words, this procedure "fills the gaps" between the estimated subject-level ideal-points by assuming that the likelihood of finding other ideal-points in the neighborhood of the estimated ones depends on their distance from them. This procedure depends on the researcher's ability to obtain reliable subject-level estimates. Since these estimates are used as direct observations from the true density function, the procedure might capitalize on the estimation errors.

Another potential application of the concepts discussed in this paper would be in the development of probabilistic unfolding models (e.g., De Soete et al. 1986). By using the multimodal extensions discussed here, one could make more realistic assumptions about the underlying distribution of ideal-points in these mapping procedures, allowing for the existence of more than one cluster of ideal-points in the attribute space.

References

- Batsell, R. R. (1980). On the use of hidden surface routines in the analysis of marketing data. *Journal of Marketing*, 44, 102-6.
- Bennet, R. S., & Hays, W. L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, 25, 27-43.
- Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in social science. Volume I* (pp. 105-155). New York: Seminar Press.
- Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice: Papers in honor of Clyde Coombs* (pp. 234-289). Bern: Hans Huber.
- Chapman, R. G., & Staelin, R. (1982). Exploiting rank order choice sets data within the stochastic utility model. *Journal of Marketing Research*, 19, 288-301.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Coombs, C. H., & Avrunim, G. C. (1977). Single-Peaked preference functions and the theory of preference. *Psychological Review*, 84, 216-30.
- Cooper, L. G., & Nakanishi, M. (1983). Two logit models for external analysis of preferences. *Psychometrika*, 48, 607-20.
- Donthu, N. (1986). *Flexible ideal point mapping and product positioning*. Unpublished doctoral Dissertation, University of Texas at Austin.
- De Soete, G. J., Carroll, J. D., & DeSarbo, W. S. (1986). The wandering ideal-point model: A probabilistic multidimensional unfolding model for paired comparisons data. *Journal of Mathematical Psychology*, 30, 28-41.
- Fornell, C., & Rust, R. T. (1989). Incorporating prior theory in covariance structure analysis: A Bayesian approach. *Psychometrika*, 54, 249-259.
- Half, H. M. (1976). Choice theories for differentially comparable alternatives, *Journal of Mathematical Psychology*, 14, 244-246.

- Hays, W. L., & Bennet, J. F. (1961). Multidimensional unfolding: Determining configuration for complete rank order preference data. *Psychometrika*, 26, 221-238.
- Judge, G. G., Griffiths, W. E., Hill, R. C., & Lee, T. (1980). *The theory and practice of econometrics*. New York: Wiley.
- Kamakura, W. A., & Srivastava, R. K. (1986). An ideal-point probabilistic choice model for heterogeneous preferences. *Marketing Science*, 5, 199-218.
- McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering*. New York: Marcel Dekker.
- Shocker, A. D., & Srinivasan, V. (1974). A consumer based methodology for the identification of new product ideas. *Management Science*, 20, 921-937.
- Srinivasan, V., & Shocker, A. D. (1973). Linear programming techniques for multidimensional analysis of preferences. *Psychometrika*, 38, 337-369.
- Tversky, A., & Russo, J. E. (1969). Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, 6, 1-12.

Manuscript received 8/2/88

Final version received 7/21/90