THE DISTRIBUTION OF UNOBSERVED VARIABLES

Robert E. Hall

Department of Economics
MIT

[redacted]

January 1974

The research reported here was carried out while the author was on a leave supported by the Social Science Research Council. Support was also provided by the National Science Foundation. Edward Leamer, ~~and~~ Gary Chamberlain, and Martin Weitzman made useful comments on an earlier version. I am especially indebted to Jerry Hausman for lengthy discussions and for pointing out the key reference, Karlin and Studden (1966).

1. Introduction

Statistical models of unobserved variables seem destined for an increasing role in econometric work. Especially in cross-sections, the differences in the values of the left-hand variables among observations with identical values of the right-hand variables are sufficiently large to justify careful analysis of the apparently random component of the behavior under study. The simple characterization of randomness implicit in the stochastic specification of the regression model seems inadequate when the right-hand variables in a problem account for only a small portion of the dispersion of the left-hand variable. In a number of social sciences, the following elaboration of the stochastic model has been found useful: Part of the apparent randomness in individual behavior arises from the distribution among the population of an unobserved characteristic, indexed by a scalar, θ . The rest is attributable to the genuine randomness of the behavior of individuals with identical values of θ . Examples of models of this kind include (i) ~~the~~ accident-proneness: θ is the parameter of a Poisson probability distribution of the number of accidents in the course of a year; (ii) mover-stayer models: θ is the probability of moving from one activity to another and is higher for some kinds of people (movers) than for others (stayers); (iii) sex composition of families: θ is the probability ~~of~~ that a particular child is a girl; some couples are girl-prone and others are boy-prone; (iv) distribution of true scores: θ is the true IQ of an individual and is measured with random error by a number of tests; the problem is to ~~di~~ find the distribution of the true score within the population given the distribution of the observed scores; and (v) the permanent income model: θ is permanent income; observed income is the sum of permanent income and a transitory component.

Several of these applications are ~~dexcr~~ described in greater detail in section 2 of this paper. They share one important limitation: They start with restrictive assumptions about the form of the distribution of the unobserved variable. The present paper takes up the following questions: What can be discovered about the underlying distribution of characteristics from the observed body of data? Are the assumptions about the distributions of unobserved characteristics made by previous authors verifiable, or must they be accepted on pure faith?

A general statistical model suitable for this discussion is the following: Let y be the observed variable and let G be its cumulative distribution:

$$G(Y, \theta) = \text{Prob}(y < Y \mid \theta)$$

The problem, then, is to derive as much information as possible about the distribution of the unobserved variable, say $F(\theta)$, given the distribution of the observed variable.

Within this general framework, two classes of models have been studied. In the first, y is assumed to be integer-valued and its distribution within a group of individuals of a given type, θ , is assumed to be of known form, for example, Poisson or binomial. The first three applications listed above have this form. In the second class, y is taken as continuous, G is not assumed to be known but does have the particular form $G(Y, \theta) = H(Y - \theta)$. Thus y can be written as the sum of θ and another random variable, u, expressing the transitory component of individual experience:

$$y = \theta + u$$

The problem in this class of models is to find not only G but also H, the distribution of the transitory component. The present paper considers only the first class of models. Analysis of the second class appears in a companion paper, Carlton and Hall (1974).

The third section of this paper embeds the statistical model in the mathematical theory of Tchebycheff systems. Various characterizations of the limits of knowledge of the distribution of the unobserved variable are offered, including an identification theorem. The fourth section discusses and applies an important generalization of the Tchebycheff inequlaity due to Daniel McFadden. The fourth section works out a simple example dealing with the sex composition of families. The paper concludes with some remarks on the mover-stayer model and possible extensions of the theory.

3. Statistical and Mathematical Theory

In general we will be concerned with the distribution of an integer-valued function, $g(y)$, of the vector, y, of observations from the same individual. For example, in the study of the sex composition of families reported briefly in Section 4, y_1 and y_2 are "dummy variables" for the sexes of the first and second children in a family (0 if a boy and 1 if a girl), and we examine $g(y)=y_1+y_2$, the number of girls in the family. In the generalized mover-stayer model, we will study the distribution of the number of spells in a given state. Throughout, we will suppose that $g(y)$ takes on the values 1 through M, and will denote its distribution as $a(\theta)$:

$$\text{Prob}\left[\, g(y)=i \;\middle|\; \theta \right] = a_i(\theta) \qquad (2.1)$$

We suppose further that θ is distributed among the population according to the cumulative distribution function $F(\bar{\theta})$ giving the fraction of the population whose unobserved characteristic is less that $\bar{\theta}$:

$$F(\bar{\theta}) = \text{Prob}\left[\, \theta < \bar{\theta} \,\right] \qquad (2.2)$$

Since we need to deal with distributions that assign positive probability to a single value of θ, we also define

$$F^{*}(\bar{\theta}) = \text{Prob}\left[\,\theta \leq \bar{\theta}\,\right] \tag{2.3}$$

The observed distribution of g(y) among the members of the population is the distribution of g(y) conditional on θ, weighted by the distribution of θ:

$$\phi_i = \text{Prob}\ \overset{g(y)}{[y = i]} = \int_0^1 a_i(\theta)dF(\theta)\ , \qquad (2.4)$$

[redacted] There is a substantial statistical literature dealing with problems of this form. In the vocabulary of that literature, equation 2.4 is a <u>mixture</u>. The distribution $a(\theta)$ is the <u>kernel</u> and $F(\theta)$ is the <u>mixing distribution</u>. A survey of the statistical theory of mixtures appears in Maritz (1970), Chapter 2. In addition, there is an important body of mathematical thought about problems of the sort considered here. In the mathematical literature, equation 2.4 is called a <u>Tchebycheff system</u> (see Karlin and Studden (1966), Chapters I through V). It appears that statistical and mathematical work in this area has proceeded almost completely independently. The mathematical theory is substantially more general and more fully developed, so it forms the basis for this paper.

Our problem is to obtain information about the distribution of the unobserved component, $F(\theta)$, given the observed probability ϕ and the known kernel $a(\theta)$. In this section we present theorems that give a fairly precise characterization of the limits of knowledge about $F(\theta)$. Most of these theorems are simply re-interpretations of results of Krein (1951) and other mathematical students of Tchebycheff systems.

We begin with the

> <u>Assumption of Distinct Types</u>: The matrix $[a(\theta_1),\ldots,a(\theta_M)]$ has rank M for any distinct set of types $\theta_1,\ldots,\theta_M$.

This assumption is the defining characteristic of a Tchebycheff system.

It rules out models where the probabilities associated with one particular type of individual can be expressed as a linear combination of the probabilities associated with M - 1 or fewer other types. This assumption does not seem unduly strong, and it is satisfied by the applications studied in this paper.

Next we define two useful constructions. First,

$$\Phi = \{\phi \text{ satisfying } \phi = \int_0^1 a(\theta)dF(\theta) \text{ for some } F(\theta)\}. \qquad (2.5)$$

Here we consider all F(θ) that are non-decreasing, continuous from the left, and have a finite number of discontinuities. Φ is the set of all possible observed probabilities consistent with a given problem as defined by a(θ). Second,

$$V(\phi) = \{F(\theta) \text{ satisfying } \int_0^1 a(\theta)dF(\theta) = \phi\}. \qquad (2.6)$$

V(φ) is the set of all distributions of unobserved types in the population that are consistent with a particular observed probability, φ. The essence of the problem is that V(φ) may contain a variety of distributions. Our characterization of the limits of knowledge about F(θ) deals, therefore, with the extremal members of V(φ).

The first theorem establishes that no observed probability proves that there are more than (M + 2)/2 different types in the population (proofs and references appear in the appendix to this section):

<u>Theorem 2.1</u>: For any $\phi \in \Phi$, there exists a cdf, $F(\theta) \in V(\phi)$, with no more than (M + 2)/2 points of increase.

If we let f_j be the mass at one of the points of increase, θ_j, then Theorem 2.1 shows that it is always possible that ϕ is a discrete mixture:

7

$$\phi = \sum_{j=1}^{N} f_j \; a(\theta_j) \qquad (2.\cancel{5})$$

with $N \leq (M + 2)/2$. Here f_j is the fraction of the population having type θ_j. This result appears independently in the statistical literature on mixtures in the form of an identification theorem: Given ϕ, one can calculate unique f_j and θ_j satisfying

8

$$\sum_{j=1}^{N} f_j \; a(\theta_j) = \phi \qquad (2.\cancel{6})$$

only if $N \leq 1 + M/2$. See Teicher (1963), p. 1269.

The second theorem shows that for any observed ϕ (with one class of exceptions) we cannot rule out the possibility that a positive fraction of the population has an arbitrary type, θ*:

> Theorem 2.2: Suppose ϕ is in the interior of Φ and suppose θ* is an arbitrary type in [0,1]. Then there is a cdf, $F(\theta)$, in $V(\phi)$ with positive mass $\rho(\theta^*)$ ~~at θ*~~. $= F^+(\theta^*) - F(\theta^*)$ at θ^*

This result imposes a limitation on the form of knowledge about F that we can deduce from ϕ: Except in borderline cases, we will never be able to state that any particular type, or any range of types, is non-existent in the population. On the other hand, $\rho(\theta^*)$ may be close to zero; the theorem does not prevent us from finding useful bounds on the fraction of the population of a certain type or range of types.

The next theorem provides a bound on the fraction of the population of type θ^*:

Theorem 2.3: Consider the problem of finding probabilities $f_1, \ldots, f_N$ and types $\theta_J, \ldots, \theta_N$ obeying

$$\sum_{j=1}^{N} f_j \; a(\theta_j) = \phi \qquad (2.\cancel{7}\,9)$$

where $\theta_1 = \theta^*$ and N, θ_2, θ_3 and J take on one of the following sets of values: If M is odd, either N = 1 + (M − 1)/2, J = 2, or N = 2 + (M − 1)/2, J = 4, $\theta_2 = 0$, $\theta_3 = 1$; if M is even, N = 1 + M/2, J = 3, and either $\theta_2 = 0$ or $\theta_2 = 1$. Then this system has a unique solution and f_1 is the maximal mass at θ^* for any $F \in V(\phi)$.

Thus the problem of finding the distribution of types that is most concentrated at θ^* is simply one of solving a system of M equations in M unknowns: N values of f_j and N − J + 1 values of θ_j. The solution is called the canonical representation of ϕ involving θ^*.

A related problem is to find bounds on the fraction of the population whose type is less than some value θ^*:

Theorem 2.4 (Markov-Krein Theorem):

$$\sum_{j \ni \theta_j < \theta^*} f_j \leq F(\theta^*) \leq \sum_{j \ni \theta_j \leq \theta^*} f_j$$

$$\text{for all } F \in V(\phi) \qquad (2.\cancel{8}\,10)$$

where f_j and θ_j are the canonical representation involving θ^*.

The Markov-Krein theorem shows that the canonical representation is extremal not only with regard to the mass at θ^* but also with regard to the mass below θ^*. The upper and lower bounds on $F(\theta^*)$ differ by precisely the maximal mass, f_1.

~~Unfortunately, t~~The mathematical theory of Tchebycheff systems does not provide bounds on the fraction of the population between two arbitrary types. We would like to be able to answer the following question: Suppose we have a pair of types θ_L and θ_H, and we let $P = F(\theta_H) - F(\theta_L)$, the fraction of the population between θ_L and θ_H. What are the largest and smallest values of P consistent with a particular ϕ? The Tchebycheff inequality answers this question for the particular case where ϕ gives the first two moments of $F(\theta)$. There is an extensive mathematical literature on generalizations of the Tchebycheff inequality (summarized in detail in Karlin and Studden (1966), chapters XII-XIV), but it does not contain any ~~results~~ computable bounds of sufficient generality for our purposes. Mathematicians have been concerned exclusively with sharp bounds on P, that is, bounds that are attained by some $F \in V(\phi)$, or at least that are approached arbitrarily closely by members of $V(\phi)$.

Before going on to our approach to the problem of bounds on the probability in an interval on the θ axis, which involves non-sharp bounds, we need to deal with the fundamental problem of identifiability. What conditions are required for it to be possible to find out anything about the fraction of the population in an interval? There has been a good deal of work on the identifiability of mixtures (see Maritz (1970), pp. 20-35), all using a strict definition of identifiability: A mixing distribution is said to be identifiable if its exact form can be deduced from the value of ϕ. Strong assumptions about $F(\theta)$ are required for identifiability. A leading result in the statistical literature has already appeared here as Theorem 2.1.

A much weaker notion of identifiability seems appropriate in this paper:

> Definition: The probability P is identifiable if there is some ϕ in the interior of Φ such that $V(\phi)$ contains no distributions with $P = 0$.

We gain information about P if we can show that it is positive, that some fraction of the population has types between θ_L and θ_H. A problem, as defined by $a(\theta)$, has an identifiable P if there is some observed outcome ϕ for which P must be positive. It is a remarkable fact that no additional assumptions are needed to ensure identifiability in a Tchebycheff system:

> Theorem 2.5: Every P is identifiable.

4. The Sex Composition of Families

The following example illustrates the nature of the information about the distribution of an unobserved component in a simple case. Suppose that we observe a large number of apparently identical families with two children, and suppose further that a fraction ϕ_1 of the families have no girls, ϕ_2 have one girl, and ϕ_3 have two girls. Each family has a probability θ that a given child will be a girl. In terms of the general model given earlier, if y is the number of girls in a family,

$$y_t = h(\theta, u_t) \quad , \; t=1,2$$

$$= u_t \text{ where u is } \text{binary} \text{ with parameter } \theta, \tag{4.1}$$

and ϕ is the distribution of y_1+y_2.

If all families have the same θ, then ϕ will be the binomial distribution:

$$\phi_1 = (1-\theta)^2; \quad \phi_2 = 2\theta(1-\theta); \quad \text{and} \quad \phi_3 = \theta^2 \tag{4.2}$$

If θ varies among families, then ϕ will be the mixed binomial,[2]

$$\phi_1 = \int_0^1 (1-\theta)^2 dF(\theta) \; ; \tag{4.3}$$

$$\phi_2 = \int_0^1 2\theta(1-\theta) dF(\theta) \; ; \tag{4.4}$$

$$\phi_3 = \int_0^1 \theta^2 dF(\theta) \; . \tag{4.5}$$

---

[2] This possibility has been discussed in the literature on mathematical demography (for example, Goodman (1961) and Weiler (1959)). This treatment of the sex composition of families is only an example and does not consider other aspects of the problem, especially the effects of the efforts of parents to influence the composition through stopping rules. On this, see Ben Porath and Welch (1972).

In our earlier notation, the kernel is:

$$a(\theta) = \begin{bmatrix} (1 - \theta)^2 \\ 2\theta(1 - \theta) \\ \theta^2 \end{bmatrix} . \tag{4.6}$$

Ben Porath and Welch (1972) report the following distribution for the sexes of the first two children of American families: $\phi_1 = 0.262$, $\phi_2 = 0.497$, and $\phi_3 = 0.241$. The mean of this distribution is 0.979, suggesting that if θ had a single value, it would be half the mean, 0.489. However, the binomial distribution with parameter $\theta = 0.489$ is [0.261, 0.500, 0.239], which has somewhat less dispersion than the observed ϕ. No single value of θ can explain the observed distribution of sexes, so we are forced to consider a distribution of the propensity to have girls, θ, within the population.

The theory of Tchebycheff systems discussed in Section 2 focuses attention on the canonical representations involving alternative values of θ^*, a preassigned type. Since M is 3 for this problem, either $N = 2$ and $J = 2$, in which case the canonical representation requires solving for f_1, f_2, and θ_2, or $N = 3$ and $J = 4$, in which case the canonical representation requires solving a linear system for f_1, f_2, and f_3. In both cases f_1 is the upper bound on the fraction of the population that has probability θ^* of having a girl, by Theorem 2.3. Further, from Theorem 2.4, when $N = 2$ and $\theta^* < \theta_2$, f_1 is the upper bound on the fraction of the population with θ less than θ^*, $F(\theta^*)$; when $N = 2$ and $\theta^* > \theta_2$, f_1 is the upper bound on the fraction of the population with θ at or above

θ^*, $1 - F(\theta^*)$, and when $N = 3$, $f_1 + f_2$ is the upper bound on $\overset{+}{F}(\theta^*)$ and $f_1 + f_3$ is the upper bound on $1 - F(\theta^*)$. Table 1 presents canonical representations for a variety of values of θ^*. For θ^* outside a short interval enclosing 0.489, the canonical representation has only one additional type, θ_2. The first part of Table 1 shows a variety of representations of this kind. When θ^* is extreme, the representation gives a low weight (f^*) to θ^* and a high weight to a θ_2 that is close to 0.489. As θ^* approaches 0.489, it receives higher weight and the second type, θ_2, becomes more extreme. At the critical points $\theta^* = 0.4868$ and $\theta^* = 0.4923$, θ_2 reaches 1, and we enter the region where the representation gives weight to three values of θ: the two extremes, $\theta = 0$ and $\theta = 1$,

Table 1

Canonical Representations for the Mixed Binomial Model of the Sex Composition of Families

θ^*	θ_2	f^*	f_2
0.0	.4923	.0058	.9942
0.30	.4968	.0372	.9628
0.40	.5050	.1479	.8521
0.48	.6358	.9390	.0610
0.50	.3571	.9265	.0735
0.60	.4769	.1022	.8978
0.70	.4829	.0304	.9696
1.00	.4868	.0053	.9947

θ^*	f_2 ($\theta = 0$)	f^*	f_3 ($\theta = 1$)
.49	.0034	.9944	.0022

and $\theta = \theta^*$. One such representation is shown in the second part of Table 1. According to this representation, the observed distribution of sex compositions could be generated by a population in which 99 44/100% of couples had a probability of 0.49 of having girls, 0.34% had nothing but boys, and 0.22% had nothing but girls.

From Table 1 we can derive the Markov-Krein bounds $\underline{P}$ and $\bar{P}$ for the fraction of the population with θ between 0 and θ_H. These bounds are presented in Table 2, along with the outside and inside bounds calculated by the methods of Section 3. All of the bounds agree that it is quite possible that no couple has a probability of having a girl below 0.48 and also possible that none has a probability above 0.50 (but

Table 2

Bounds on the Fraction of the Population with θ between 0 and θ_H

θ_H	$\underline{P}_0$	$\underline{P}$	$\underline{P}_I$	$\bar{P}_I$	$\bar{P}$	$\bar{P}_0$
0.30	0	0	0	~~.0236~~ .0353	.0372	.0375
0.40	0	0	0	~~.1018~~ .1344	.1479	~~.1500~~ .1499
0.48	0	0	0	~~.7545~~ .8686	.9390	~~.9543~~ .9413
0.49	~~.0033~~ 0	.0034	~~.0038~~ .0034	~~.9853~~ .9856	~~.9966~~ .9978	.9979
0.50	~~.0530~~ .0735	.0735	~~.0742~~ .0736	1.0	1.0	1.0
0.60	~~.8720~~ .8757	.8978	~~.8930~~ .8982	1.0	1.0	1.0
0.70	~~.9657~~ .9651	.9696	.9697	1.0	1.0	1.0

Explanation:

$\underline{P}$, $\bar{P}$: Markov-Krein exact bounds, derived from Table 1.

$\underline{P}_0$, $\bar{P}_0$: Outside bounds, with 201 equally spaced ~~endpoints 0, .15, .25, .30, .35, .38, .40, .42, .47, .48, .485, .49, .495, .50, .52, .58, .60, .62, .65, .70, .75, .85, 1.00. S_0 has 2776 vertices~~ end points.

$\underline{P}_I$, $\bar{P}_I$: Inside bounds; same points as above. ~~S_I has 368 vertices.~~

there must be some couples with probabilities either below 0.48 or above 0.50). There may be as many as 3.72% with probabilities below 0.30, as many as 14.79% below 0.40, and as many as 93.90% below 0.48. At least 0.34% and possibly as much as 99.66% of the population have probabilities below 0.49. At least 7.35% have probabilities below 0.50, at least 89.78% below 0.60, and at least 96.96% below 0.70. [redacted] Both outside bounds ($\underline{P}_O$ and $\bar{P}_O$) and the lower inside bound $\underline{P}_I$ perform well as approximations to the sharp bounds. [redacted] $\bar{P}_I$ is [redacted] somewhat too small, especially at 0.48. Recalculation of the inside bounds with a finer partition would remedy this problem.

Table 3 presents bounds for various intervals that do not begin at zero. No Markov-Krein sharp bounds are available for these intervals,

Table 3

Bounds on the Fraction of the Population with θ between θ_L and θ_H

θ_L	θ_H	$\underline{P}_O$	$\underline{P}_I$	$\bar{P}_I$	$\bar{P}_O$
0.30	0.70	~~.9625~~ .9623	~~.9697~~ .9647	1.0000	1.0000
0.40	0.60	~~.8500~~ .8452	~~.8851~~ .8621	1.0000	1.0000
0.40	0.50	0	0	~~.9955~~ .9970	~~1.0000~~ .9987
0.40	0.48	0	0	~~.7545~~ .8686	~~.9543~~ .9413
0.48	0.50	0	0	~~.9944~~ .9949	.9954
0.48	0.49	0	0	~~.9853~~ .9856	~~.9953~~ .9953
0.49	0.50	0	0	~~.9943~~ .9944	.9944
0.50	0.70	0	0	~~.9258~~ .9265	~~.9470~~ .9299

For explanation, see Table 2.

so the only way to judge the sharpness of the outside bounds is through the inside bounds. At least 96.2~~5~~3% of all couples have probabilities between 0.30 and 0.70, and we know that there exists a distribution consistent with ϕ in which only 96.~~0~~47% of the population lies between 0.30 and 0.70. On the other hand, it is possible that 99.54% of the population has θ between 0.48 and 0.50, and we know for sure that ~~99 44/100~~ 99.49 % can be in this interval.

A fairly wide variety of distributions of the propensity to have girls is consistent with the observed data on the distribution of the number of girls among the first two children. Although little can be done to localize the distribution in the vicinity of 0.5, our methods give fairly specific information about the fraction of the population with extreme propensities. The data are not consistent with any distributions with large fractions of the population having extreme values of θ. An increase in the number of times each unit is observed, in this case the number of children, would refine our knowledge considerably. A study of sex composition that examined more than the first two children would need to deal explicitly with the problem of stopping rules, however.

6 ~~5~~. Mixed Markov Processes and the Mover-Stayer Model

This section illustrates the application of the methods discussed earlier to a problem of considerable interest in the study of social mobility. Suppose there are two states that an individual may occupy in each period: poor or not poor, employed or not employed, lower class or middle class, or some other dichotomy. Suppose further that a Markov process governs transitions between the states; there is a probability θ that an individual in the first state in one period will move to the second state in the next period, and a probability δ that an individual in the second will move to the first. The probabilities of remaining in the states are then $1 - \theta$ and $1 - \delta$ respectively. Models of this kind fitted to data on observed transitions of individuals under the assumption that θ and δ are the same for all of them have suffered from an important defect: They understate the probability that an individual will remain for many successive periods in the same state, even though they predict correctly the probability that an individual chosen at random from the inhabitants of one state will move to the other state in the next period (Blumen, Kogan, and McCarthy (1955)).

The mover-stayer model resolves this paradox by assuming that there are actually two kinds of people, movers, who have positive θ, and stayers, whose θs are zero. The probabilities of observed transitions are the mixture of two different Markov processes. Methods for estimating the parameters of the two processes and the single mixing probability have been developed by Goodman (196~~0~~1). Recently Spilerman (1972) has proposed an extension of the model in which the observed probabilities are treated as the mixture of all of the powers of a particular transition matrix. None of the literature on the mover-stayer model takes advantage of the statistical theory of mixtures, however.

A natural generalization of the mover-stayer model is the mixture of all Markov processes. To keep within the confines of the theory developed in this paper, however, we will suppose that individuals differ only with respect to their probability of upward mobility, θ, and that δ is known and constant within the population. Then it is appropriate to study the distribution of the number of spells in the second state over a certain number of periods, T. Individuals with high values of θ will tend to have more spells than do those with low θ. We define the observed probability, ϕ, in the following way:

$$\begin{aligned} \phi_1 &= \text{fraction of the population with no spells} \\ &\vdots \\ \phi_i &= \text{fraction with } i - 1 \text{ spells} \\ &\vdots \\ \phi_M &= \text{fraction with } M - 1 \text{ or more spells} \end{aligned}$$

Data on spells of unemployment during a year are reported by the U. S. Census Bureau in precisely this form, with $M = 4$.

We define $a_i(\theta)$ as the probability of $i - 1$ spells in T periods induced by a Markov process with parameters θ and δ. There is no simple closed form for $a_i(\theta)$, but it can be calculated from a simple ~~the following~~ recursion. Let $Q(t,i,j)$ be the probability of having $i - 1$ spells in t periods and of finishing in state j at time t. Then

$$\begin{aligned} Q(t+1,i,1) &= (1 - \theta)Q(t,i,1) + \theta Q(t,i,2) \\ Q(t+1,i,2) &= \delta Q(t,i-1,1) + (1 - \delta)Q(t,i,2) \end{aligned} \qquad (6.1)$$

with

$$\begin{aligned} Q(0,i,j) &= 0 \quad \text{if } i \neq 1 \\ Q(0,1,1) &= p^* \\ Q(0,1,2) &= 1 - p^* \\ Q(t,-1,1) &= 0 \ , \ t = 1,\ldots,T \end{aligned} \qquad (6.2)$$

Here p^* is the probability of being in the first state at time 0 and might reasonably be taken as the steady-state probability of being in the first state:

$$p^* = \frac{\delta}{\theta + \delta} \tag{5.3}$$

Finally,

$$a_i(\theta) = Q(T,i,1) + Q(T,i,2) \quad , \; i = 1,\ldots,M - 1 \tag{5.4}$$

$$a_M(\theta) = 1 - a_1(\theta) - \ldots - a_{M-1}(\theta) \tag{5.5}$$

This puts the mixture into our standard form,

$$\phi = \int_0^1 a(\theta)dF(\theta) \; . \tag{5.6}$$

All of our earlier techniques can be applied to obtain information about the distribution of the probability of upward mobility among the population. The mover-stayer model is the special case where $F(\theta)$ concentrates all its probability at $\theta = 0$ and at one other value of θ. From Theorem 2.1, if our data distinguish only among no spells, one spell, and two or more ($M = 3$), then there is always a simple mover-stayer model that explains the observed ϕ, namely the canonical representation involving $\theta^* = 0$. Other distributions will also be consistent with ϕ, however, and if the data on the number of spells are richer, the simple mover-stayer model will not generally be able to explain ϕ. In any case, the assumption that there are exactly two types of people is a highly restrictive one; our methods provide a workable method for relaxing it.

7 ~~6~~. Extensions

Many investigators are likely to be willing to make restrictive assumptions about the form of the distribution of the unobserved component in order to tighten the results by ruling out implausible distributions. This can be done through the conventional device of confining the distribution to a family indexed by a limited number of parameters. If the number of parameters is equal to the number of observed probabilities, then it is often straightforward to calculate $F(\theta)$ from ϕ. For example, if $a(\theta)$ is binomial and $F(\theta)$ is a beta distribution, then the parameters can be calculated directly from ϕ; see Maritz (1970), pp. 22-23. On the other hand, a weak parametrization that imposes nothing more than smoothness on $F(\theta)$ will usually have more than M parameters, so more than one member of the parametric family of distributions will be consistent with the observed ϕ. The problem then is essentially similar to the problem treated in this paper. In particular, if the family is linear in its parameters, the set of parameters consistent with ϕ is mathematically the same as the set S_I derived in Section 3. The family of distributions whose densities are step functions is an important example of such a family.

Second, in practice we do not observe the probabilities ϕ but only the corresponding frequencies, say $\hat{\phi}$. If we apply our methods to $\hat{\phi}$, then our bounds become random variables that estimate the bounds but are not truly bounds themselves. A confidence region enclosing $\hat{\phi}$ induces a confidence interval for each bound. The only serious problem in dealing with $\hat{\phi}$ arises when it does not lie in Φ. For example, in a small population it is possible that every family has one girl and one boy, but there is no mixture of binomial distributions that gives rise to the corresponding set of probabilies. Fortunately, if ϕ is in the interior of Φ, the probability that $\hat{\phi}$ lies outside Φ approaches zero as the sample size increases.

Third, in many applications the probabilities of alternative outcomes depend on the observed characteristics of the individual as well as on his unobserved type. The easy way to incorporate this dependence in our model is to let $F(\theta;x)$ be the distribution of θ within the subpopulation of individuals with characteristics x. Then the observed mixture also depends on x:

$$\phi(x) = \int_0^1 a(\theta)dF(\theta;x) \qquad (\cancel{6}\,7.1)$$

Given $\phi(x)$ for a particular x, we can then apply our methods to derive information about $F(\theta;x)$. In practice, we specify $\phi(x)$ as a multinomial probability depending on x in a reasonably flexible way, using a multinomial logit or other convenient specification. Note that $\phi(x)$ does not have the same structure as $a(\theta)$--for example, the study of mixed Markov processes does not involve the estimation of the parameters of a Markov process. From $\phi(x)$, we calculate bounds on $F(\theta;x)$ for representative values of x.

Fourth, the model of equation 1.2 and the methods of Section 3 (but not the Tchebycheff theory of Section 2) generalize in an obvious way to the case of several unobserved variables. Of course, problems of higher demension require correspondingly more observations from each individual in order to obtain useful knowledge of the joint distribution of the unobserved variable.

Concluding Remarks

Unobserved differences among individuals are an important source of diversity in their observed behavior. For the case in which the probability distribution among the alternatives is a known function of the unobserved type, this paper has shown that exact but not complete knowledge of the distribution can be obtained. The assumptions of previous authors about these distributions can, in fact, be tested.

References

Tom Apostol, (1957) Mathematical Analysis, Reading, Massachusetts, Addison-Wesley.

Yoram Ben Porath and Finis Welch, (1972) "Uncertain Quality: Sex of Children and Family Size," Rand Publication R-1117-NIH/RF, Aug. 30, 1972, Revised October 1972.

Isadore Blumen, Marvin Kogan, and Philip J. McCarthy, (1955) The Industrial Mobility of Labor as a Probability Process, Cornell University.

Thomas Domencich and Daniel McFadden, (1974) Development of Behavioral Models of Urban Travel Demand, North-Holland, forthcoming.

Arthur S. Goldberger (1972)"Structural Equation Methods in the Social Sciences" Econometrica, Vol. 40, pp. 979-1001, November

Leo Goodman, (1961) "Statistical Methods for the Mover-Stayer Model," Journal of the American Statistical Association, Vol. 56, pp. 841-868.

Leo Goodman, (1961) "Some Possible Effects of Birth Control on the Human Sex Ratio," Annals of Human Genetics, Vol. 25, pp. 75-81.

Zvi Griliches, (1974) "Errors in Variables and Other Unobservables" forthcoming, Econometrica

Zvi Griliches and Gary Chamberlain (1973) "Returns to Schooling of Brothers and Ability as an Unobservable Component", forthcoming

Christopher Jencks, et al., (1972) Inequality: A Reassessment of the Effect of Family and Schooling in America, New York, Basic Books.

Samuel Karlin and William J. Studden, (1966) Tchebycheff Systems: with Applications in Analysis and Statistics, New York, Wiley.

M. G. Krein, (1951) "The Ideas of P.L. Cebysev and A.A. Markov in the Theory of Limiting Values of Integrals and Their Further Developments," American Mathematical Society Translations, Series 2, Vol. 12, pp. 1-122.

J.S. Maritz, (1970) Empirical Bayes Methods, London, Methuen.

Seymour Spilerman, (1972) "Extensions of the Mover-Stayer Model," American Journal of Sociology, Vol. 78, Nov. 1972, pp. 599-626.

Henry Teicher, (1963) "Identifiability of Finite Mixtures," Annals of Mathematical Statistics, Vol. 34, pp. 1265-1269.

H. Weiler, (1959) "Sex Ratio and Birth Control," American Journal of Sociology, Vol. 65, pp. 289-299.

Appendix to Section 2

General remark. Proofs of the results in this section are all taken from Karlin and Studden (1966) (hereafter K & S). They deal with a somewhat more general problem in which $F(\theta)$ is not required to obey $\int_0^1 dF(\theta) = 1$ and $a(\theta)$ is not required to satisfy $\Sigma a_i(\theta) = 1$. In their exposition, Φ is a convex cone, while under our assumptions it is a convex subset of the unit simplex. However, the results invoked here apply without modification, because our Φ is simply the intersection of their Φ and the unit simplex.

Proof of Theorem 2.1: If ϕ is on the boundary of Φ, apply Theorem II.2.1, K & S. Otherwise, apply their Corollary II.3.1. If M is odd, $N = (M + 1)/2$.

Proof of Theorem 2.2: The appropriate cdf, $F(\theta)$, can be taken as defined in Theorem 2.3. K & S, Theorem II.3.1, establish that the mass is positive.

Proof of Theorem 2.3: K & S, Theorem II.4.1 (attributed to Krein (1951)), show that the canonical representation involving θ^* assigns maximal mass to θ^*. Existence and uniqueness of the canonical representation follow from their Theorem II.3.1 and Corollary II.3.2, respectively.

Proof of Theorem 2.4: See K & S, Theorem III.2.1.

Proof of Theorem 2.5: We need to exhibit a ϕ such that all $F \in V(\phi)$ have positive mass in the interval $[\theta_L, \theta_H]$. Define $\theta_k = \frac{N-k+1}{N}\theta_L + \frac{k-1}{N}\theta_H$. If M is odd, let $N = (M + 1)/2$ and

$$\phi = \frac{1}{N}\sum_{k=1}^{N} a(\theta_k); \tag{A2.1}$$

otherwise, let N = (M + 2)/2 and

$$\phi = \frac{1}{N+1} \sum_{k=1}^{N} a(\theta_k)$$

$$+ \frac{1}{N+1} a(0) \tag{A2.2}$$

For this ϕ, the values of θ_k and $f_k = \frac{1}{N}$ or $\frac{1}{N+1}$ are a canonical representation. By K & S's Lemma II.3.1, every $F(\theta) \in V(\phi)$ assigns positive mass to $[\theta_{k-1},\theta_k]$, so clearly P must be positive. Finally, K & S's Theorem II.2.1 establishes that ϕ is in the interior of Φ.