An Individualistic Approach to Item Analysis

Georg Rasch *University of Copenhagen*

1. Introduction

Traditionally the properties of a psychological test are defined in terms of variations within some specified population. In practice such populations may be selected in various reasonable ways, and accordingly the properties referred to—for example, the reliability coefficient—are not specific to the test itself but may vary according to how the population is defined. Similarly, the evaluation of a subject is usually linked up with a population by a standardization of some kind and is therefore not specific to the subject per se. Our aim is to develop probabilistic models in the application of which the population can be ignored. It was a discovery of some mathematical significance that such models could be constructed, and it seemed remarkable that data collected in routine psychological testing could be fairly well represented by such models.

In a previous study (2) an attempt was made to build up a general framework within which the models of an earlier study (3) appear to be special cases, and some properties of this general framework were recognized. But only recently it has become quite clear that this model is in fact *the complete answer to the requirement that statements about the parameters and adequacy of a discrete probabilistic model be objective in a sense to be fully specified.*

At present, at least, the theory leading to this result is rather involved, and it is not going to be a main topic for this paper. However, it is intended that the following discussion on one of the models in the earlier study (3), the model for item analysis in case of only two possible answers, should demonstrate the nature of the type of objectivity we are aiming at, thus pointing to the more general problem to be treated elsewhere.

2. Data

The kind of situation to be considered is the following, in which a fairly large number of subjects were given an intelligence test. Two subtests, N (completing numerical sequences) and F (analyzing geometric figures), are of particular interest. The time allowed for N was chosen so that very few of the subjects could be expected to achieve an appreciably larger num-

Excerpt from READINGS IN MATHEMATICAL SOCIAL SCIENCE (1966)

ber of correct answers even with unlimited time. (This fact was ascertained by independent experimental evidence.) Therefore items that were not answered were counted as incorrect, and one of the two responses, correct (+) or incorrect (−), was recorded for each item. Test F was scored in a similar way.

3. Model

The model to be suggested is based on three assumptions:

1. To each situation of a subject (ν) having to solve a test item (i), there corresponds a probability of a correct answer which we will write in the form

(3.1) $p\{+\mid \nu, i\} = \dfrac{\lambda_{\nu i}}{1+\lambda_{\nu i}},\ \lambda_{\nu i} \geq 0.$

2. The situational parameter $\lambda_{\nu i}$ is the product of two factors,

(3.2) $\lambda_{\nu i} = \xi_\nu \epsilon_i,$

ξ_ν pertaining to the subject, ϵ_i to the item.
3. All answers, given the parameters, are stochastically independent.

Each of these assumptions calls for some comments.

1. For description of observations two apparently antagonistic types of models are available, deterministic models (such as the law of gravitation) and stochastic models (such as Mendel's laws of heredity). However, the choice of one type or the other does not imply that the phenomena observed *were* causally determined or that they *did* occur by chance.

Even if it were believed that certain phenomena could be "explained causally" (whatever such a phrase may mean), a stochastic model may be preferable (as in thermodynamics).

Although adopting a probabilistic model for describing responses to an intelligence test, we have taken no sides in a possible argument about responses being ultimately explainable in causal terms.

2. In many psychophysical threshold experiments a subject is exposed to the same stimulus a large number of times. Assuming that the repetitions do not affect the judgments of the subject, this procedure gives the opportunity of estimating each $\lambda_{\nu i}$ separately and hence of studying directly how the situational parameter varies with subject and with strength of stimulus. In such a situation we may or may not observe the multiplicative rule laid down in 3.2.

For the intelligence tests we shall deal with, experience has shown that on one repetition the results are usually somewhat improved. A large number of repetitions have not been tried, mainly because the questions are such that it seems almost certain that several of them will easily be recognized after a few repetitions. Therefore the possibility of a direct approach to the estimation of response probabilities seems remote.

To compensate, we have recourse to an assumption that may seem rather bold, possibly even artificial, namely that $\lambda_{\nu i}$ can be factored into a subject parameter and an item parameter. This, combined with the two other assumptions, produces a model that turns out to have rather remarkable properties, some of which even lead to the possibility of examining how well the model represents the data (see Section 6).

Provided the two kinds of parameters can be operationally defined, they also have a clear meaning, as seen by inserting 3.2 into 3.1:

(3.3) $$p\{+\mid \nu, i\} = \frac{\xi_\nu \epsilon_i}{1+\xi_\nu \epsilon_i}.$$

Thus, if the same person is given items with ϵ_i approaching 0, then his probability of giving a correct answer approaches 0 while his probability of giving an incorrect answer tends toward unity. And that is true for every person, provided the model holds. Similarly, when ϵ_i gets large, the probability of + tends toward 1 and the probability of − toward 0. Since with increasing ϵ_i the items become easier, we may call ϵ_i *the degree of easiness* of item i.

On the other hand, giving the same item to persons with ξ_ν approaching 0, we get probabilities of correct answers tending toward 0, while if ξ_ν increases indefinitely the probability tends toward 1. This holds for every item. Thus we may colloquially call ξ_ν *the ability of subject* ν with respect to the kind of items in question.

In the definition of ξ_ν and ϵ_i there is an inherent indeterminacy. In fact, if $\xi_\nu, \nu = 1, \ldots, N$ and $\epsilon_i, i = 1, \ldots, k$ is a set of solutions to the equations

(3.4) $$\xi_\nu \epsilon_i = \lambda_{\nu i}$$

then, if ξ'_ν, ϵ'_i is another set of solutions, the relation

(3.5) $$\xi'_\nu \epsilon'_i = \xi_\nu \epsilon_i$$

must hold for any combination of ν and i. Thus

(3.6) $$\frac{\xi'_\nu}{\xi_\nu} = \frac{\epsilon_i}{\epsilon'_i}$$

must be a constant, say, and accordingly the general solution is

$$(3.7)\qquad \xi'_\nu = \alpha\xi_\nu,\ \epsilon'_i = \frac{1}{\alpha}\epsilon_i,\ \alpha > 0 \text{ arbitrary.}$$

The indeterminacy can be removed by the choice of one of the items, say $i = 0$, as the *standard item* having "a unit of easiness," in multiples of which the degrees of easinesses of the other items are expressed.

By this choice, or an equivalent one, the whole set of ξ'_ν's and ϵ'_i's is fixed. In particular

$$(3.8)\qquad \xi_\nu = \lambda_{\nu 0},$$

that is, the parameter of a subject is a very simple function of *his probability of giving a correct answer to the standard item,*

$$(3.9)\qquad \lambda_{\nu 0} = \frac{p\{+\mid \nu,0\}}{1-p\{+\mid \nu,0\}}$$

being the "betting odds" on a correct answer. Now we may be able to find a person who has in fact his $\xi = 1$. We may refer to him as a *standard subject* ($\nu = 0$). And then *the item parameter*

$$(3.10)\qquad \epsilon_i = \lambda_{0i}$$

is the same simple function of *the probability that the standard person gives a correct answer to this item.*

3. To some psychologists the assumption of stochastic independence at first sight appears to be rather startling, since it is well known that usually quite high correlation coefficients between responses to different items are found.

Correlated items are, however, a consequence of the assumption. With moderate variation of ξ, say from 0.1 to 10, we will obtain quite high correlation coefficients. But, of course, if ξ is the same, or nearly the same, for all individuals, the correlations become zero, or nearly zero. Under this model the interitem correlations do not represent intrinsic properties of the items, but are mainly determined by the variations in the person parameters.

Let $p\{\binom{i}{+}\}$ and $p\{\binom{i}{-}\}$ stand for a person's probabilities of a positive and negative response, respectively, to item i. Considering next his possible responses to two items, i and j, they can also be allotted probabilities: $p\{\binom{i}{+}, \binom{j}{+}\}$, and so on. Now our third assumption states, among other things, that his responses to i and j should be "stochastically independent." Technically this is expressed in the following relations:

$$(3.11)\qquad \begin{cases} p\{\binom{i}{+}, \binom{j}{+}\} = p\{\binom{i}{+}\}\,p\{\binom{j}{+}\} \\ p\{\binom{i}{+}, \binom{j}{-}\} = p\{\binom{i}{+}\}\,p\{\binom{j}{-}\} \text{ etc.} \end{cases}$$

If in the first of these equations we divide by $p\{\binom{j}{+}\}$ and in the second by $p\{\binom{j}{-}\}$, we get

$$(3.12)\qquad p\{\binom{i}{+}\} = \frac{p\{\binom{i}{+}, \binom{j}{+}\}}{p\{\binom{j}{+}\}} = \frac{p\{\binom{i}{+}, \binom{j}{-}\}}{p\{\binom{j}{-}\}}.$$

The ratio (3.12) of the two probabilities is *the conditional probability of a + answer to* i, *given a + answer to* j. The notation is

$$(3.13)\qquad p\{\binom{i}{+} \mid \binom{j}{+}\} = \frac{p\{\binom{i}{+}, \binom{j}{+}\}}{p\{\binom{j}{+}\}}.$$

Thus the relations (3.12) can be written

$$(3.14)\qquad p\{\binom{i}{+} \mid \binom{j}{+}\} = p\{\binom{i}{+} \mid \binom{j}{-}\} = p\{\binom{i}{+}\},$$

that is, *the probability of a plus answer to* i *is independent of whether the answer to* j *is + or −*; it is just the probability of a plus answer to i.

And of course the same holds for a minus answer to i. This is a specification of the statement that the answers to i and j are *stochastically independent.*

Assumption 3 also requires that for each subject the answers to all questions be stochastically independent. Technically this is expressed in the equation

$$(3.15)\qquad p\{\binom{1}{+},\binom{2}{+},\ldots\binom{k}{+}\} = p\{\binom{1}{+}\}\,p\{\binom{2}{+}\}\ldots p\{\binom{k}{+}\}$$

and all its analogues. The content of this statement is that *the probability of a certain answer to an item or of a combination of answers to a set of items is unaffected by the answers given to the other items.*

4. Comparison of two items

As an introduction to the more general treatment of the model in Section 5 we will consider how two items can be compared.

According to 3.11 and 3.3, the probability of correct answers to both item i and item j is

(4.1)
$$\begin{cases} p\{\binom{i}{+},\binom{j}{+} \mid \xi\} = p\{\binom{i}{+} \mid \xi\}p\{\binom{j}{+} \mid \xi\} \\ \qquad = \dfrac{\xi^2\epsilon_i\epsilon_j}{(1+\xi\epsilon_i)(1+\xi\epsilon_j)} \end{cases}$$

for a subject with the parameter ξ. Similarly,

(4.2)
$$p\{\binom{i}{+},\binom{j}{-} \mid \xi\} = \frac{\xi\epsilon_i}{(1+\xi\epsilon_i)(1+\xi\epsilon_j)},$$

(4.3)
$$p\{\binom{i}{-},\binom{j}{+} \mid \xi\} = \frac{\xi\epsilon_j}{(1+\xi\epsilon_i)(1+\xi\epsilon_j)},$$

(4.4)
$$p\{\binom{i}{-},\binom{j}{-} \mid \xi\} = \frac{1}{(1+\xi\epsilon_i)(1+\xi\epsilon_j)}.$$

With the notations

(4.5)
$$a_i = \begin{cases} 1 \text{ in case of answer } + \text{ to item } i \\ 0 \text{ in case of answer } - \text{ to item } i \end{cases}$$

and

(4.6)
$$a_. = a_i + a_j,$$

the probabilities of $a_. = 0$ and 2 are given by 4.1 and 4.4, while the probability of $a_. = 1$ is the sum of 4.2 and 4.3:

(4.7)
$$p\{a_. = 1 \mid \xi\} = \frac{\xi(\epsilon_i+\epsilon_j)}{(1+\xi\epsilon_i)(1+\xi\epsilon_j)}.$$

Now the conditional probability of $a_i = 1$ provided $a_. = 1$ is—analogous to 3.13—obtained by dividing 4.7 into 4.2. However, by that operation the common denominator and ξ in the numerators cancel and we are left with

(4.8)
$$p\{a_i = 1 \mid a_. = 1, \xi\} = \frac{\epsilon_i}{\epsilon_i+\epsilon_j},$$

irrespective of the subject parameter ξ.

Considering, then, a number, n, of subjects, all of whom happened to have $a_. = 1$, the probability that c of them have $a_i = 1$ (and thus $a_j = 0$) is given by the binomial law:

(4.9)
$$p\{c \mid n\} = \binom{n}{c}\left(\frac{\epsilon_i}{\epsilon_i+\epsilon_j}\right)^c\left(\frac{\epsilon_j}{\epsilon_i+\epsilon_j}\right)^{n-c}.$$

Accordingly, by the relation

(4.10)
$$\frac{\epsilon_i}{\epsilon_i+\epsilon_j} \approx \frac{c}{n}$$

the ratio (ϵ_i/ϵ_j) is estimated *independently of the subject parameters*, the distribution of which is therefore irrelevant in this connection.

Furthermore, we may get a check on the model by first stratifying the subjects according to any principle—educational level or socioeconomic status or even randomly—and then applying 4.10 to each of the groups. For the model to hold, the ratio ϵ_i/ϵ_j should be the same in all of the groups and the variation of the estimates obtained should therefore concur with the binomial distributions 4.9.

The appropriate test for this constancy has a remarkable property. Denote the within-group c's and n's by c_g, n_g, with $g = 1, \ldots, h$, and their totals by $c_.$ and $n_.$. Since the groups could be collected into one group of size $n_.$ to which 4.9 applies, we have

(4.11)
$$p\{c_. \mid n_.\} = \binom{n_.}{c_.}\left(\frac{\epsilon_i}{\epsilon_i+\epsilon_j}\right)^{c_.}\left(\frac{\epsilon_j}{\epsilon_i+\epsilon_j}\right)^{n_.-c_.}.$$

On the other hand, the joint probability of the numbers $c_1, \ldots, c_h$, due to their stochastic independence, is

(4.12)
$$p\{c_1, \ldots, c_h \mid n_1, \ldots, n_h\} = \prod_{g=1}^{h} \binom{n_g}{c_g} \cdot \left(\frac{\epsilon_i}{\epsilon_i+\epsilon_j}\right)^{c_g}\left(\frac{\epsilon_j}{\epsilon_i+\epsilon_j}\right)^{n_g-c_g}$$

In consequence the conditional probability of $c_1, \ldots, c_h$ given the total $c_.$, obtained by dividing 4.11 into 4.12, becomes independent of ϵ_i and ϵ_j:

(4.13)
$$p\{c_1, \ldots, c_h \mid c_., n_1, \ldots, n_h\} = \frac{\prod_{g=1}^{h}\binom{n_g}{c_g}}{\binom{n_.}{c_.}}.$$

It follows that as far as the items i and j are concerned, *the testing of the model* can be carried out in a way that is *independent of all of the parameters*.

In the formal derivation of the fundamental relation 4.8, subjects and items can of course be interchanged. Thus the comparison of two subjects μ and ν by means of a single item with parameter ϵ leads to the conditional probability

(4.14)
$$p\{a_\mu = 1 \mid a_. = 1, \epsilon\} = \frac{\xi_\mu}{\xi_\mu+\xi_\nu},$$

where a_μ, a_ν, and $a_.$ have a meaning similar to 4.5 and 4.6. Probability 4.14 is *independent of which item was used.*

In principle, therefore, it should be possible to estimate the ratio independently of the item parameters. In practice, however, this method does not work, because the number of items—in contrast to the number of subjects—usually is small.

5. Generalization to k items

In generalizing the results of the preceding section we will first consider the responses of an individual with parameter ξ to k items. With the notation 4.5 and the adaptation

$$a_. = a_1 + \ldots + a_k \tag{5.1}$$

of 4.5, we may condense 3.3 to

$$p\{a_i \mid \xi\} = \frac{(\xi\epsilon_i)^{a_i}}{1+\xi\epsilon_i} \tag{5.2}$$

and the generalization of 4.1 through 4.4 to

$$\left\{\begin{aligned} p\{a_1, \ldots, a_k \mid \xi\} &= p\{a_1 \mid \xi\} \ldots p\{a_k \mid \xi\} \\ &= \frac{\xi^{a.}\ \epsilon_1^{a_1} \ldots \epsilon_k^{a_k}}{\prod_{i=1}^{k} (1+\xi\epsilon_i)}, \end{aligned}\right. \tag{5.3}$$

recalling that a_i is either zero or one. From this result we derive the probability that $a_.$ takes on a specified value r. If $r = 0$, every $a_i = 0$, and thus

$$p\{a_. = 0 \mid \xi\} = \frac{1}{\gamma(\xi)}, \tag{5.4}$$

where for short we write

$$\prod_{i=1}^{k} (1+\xi\epsilon_i) = \gamma(\xi). \tag{5.5}$$

We can obtain $r = 1$ in k different ways—

$$\begin{aligned} &a_1 = 1,\ a_2 = \ldots = a_k = 0, \\ &a_1 = 0,\ a_2 = 1,\ a_3 = \ldots = a_k = 0 \\ &\text{— — — — — — — — — — — —} \\ &a_1 = a_2 = \ldots = a_{k-1} = 0,\ a_k = 1 \end{aligned} \tag{5.6}$$

—with the probabilities

$$\frac{\xi\epsilon_1}{\gamma(\xi)}, \frac{\xi\epsilon_2}{\gamma(\xi)}, \ldots, \frac{\xi\epsilon_k}{\gamma(\xi)}, \tag{5.7}$$

the sum of which is the probability

$$p\{a_. = 1 \mid \xi\} = \frac{\xi(\epsilon_1 + \ldots + \epsilon_k)}{\gamma(\xi)}. \tag{5.8}$$

We can obtain $r = 2$ in $\binom{k}{2}$ different ways, namely by taking any two of the a_i's to be 1, the rest of them being 0. The probabilities of these combinations are

$$\frac{\xi^2\epsilon_1\epsilon_2}{\gamma(\xi)}, \frac{\xi^2\epsilon_1\epsilon_3}{\gamma(\xi)}, \frac{\xi^2\epsilon_2\epsilon_3}{\gamma(\xi)}, \ldots, \frac{\xi^2\epsilon_{k-1}\epsilon_k}{\gamma(\xi)}, \tag{5.9}$$

and the sum of them is

$$p\{a_. = 2 \mid \xi\} = \frac{\xi^2(\epsilon_1\epsilon_2 + \ldots + \epsilon_{k-1}\epsilon_k)}{\gamma(\xi)}. \tag{5.10}$$

In general $a_. = r$ can be obtained in $\binom{k}{r}$ different ways, by taking any r out of the k a_i's to be 1, the rest of them being 0. The probabilities of these combinations being

$$\frac{\xi^r\ \epsilon_1 \ldots \epsilon_r}{\gamma(\xi)}, \frac{\xi^r\ \epsilon_1 \ldots \epsilon_{r-1}\epsilon_{r+1}}{\gamma(\xi)}, \ldots, \frac{\xi^r\ \epsilon_{k-r+1} \ldots \epsilon_k}{\gamma(\xi)}, \tag{5.11}$$

the probability of $a_. = r$ becomes

$$p\{a_. = r \mid \xi\} = \frac{\gamma_r \xi^r}{\gamma(\xi)} \tag{5.12}$$

where, for short,

(5.13) $$\gamma_r = (\epsilon_1 \ldots \epsilon_r) + \ldots + (\epsilon_{k-r+1} \ldots \epsilon_k).$$

In particular for $r = k$, 5.13 contains only one term,

(5.14) $$\gamma_k = \epsilon_1\epsilon_2 \ldots \epsilon_k.$$

If in 5.12 we let r pass through the values $0,1, \ldots, k$, all possibilities have been exhausted and therefore the probabilities must add up to unity:

(5.15) $$\sum_{r=0}^{k} p\{a_. = r \mid \xi\} = 1.$$

Hence

(5.16) $$\gamma(\xi) = \sum_{r=0}^{k} \gamma_r \xi^r,$$

that is, γ_r are the coefficients in the expansion of the product 5.5 in powers of ξ.* If the ϵ's were known, the γ_r's could be computed and it would be possible from an observed $a_.$ to estimate ξ and to indicate the precision of the estimate—for example, in terms of confidence intervals. Thus $a_.$ is what is called an *estimator* for ξ. How to compute an estimate from the estimator is not our concern at present, but as an estimator $a_.$ has an important property.

On dividing 5.12 into 5.3 we obtain the conditional probability of the a_i's, given that their sum is r. Through this operation both the common denominator and the common power of ξ cancel and we get

(5.17) $$p\{a_1, \ldots, a_k \mid a_. = r, \xi\} = \frac{\epsilon_1^{a_1} \ldots \epsilon_k^{a_k}}{\gamma_r},$$

which is *independent of* ξ, the parameter to be estimated.

In order to realize the significance of this result we can turn to an obvious but fundamental principle of science, namely, that *if we want to know something about a quantity*—for example, a parameter of a model—*then we have to observe something that depends on that quantity*, something that changes if the quantity varies materially. For the purpose of estimating the parameter ξ of a person, the observations $a_1, \ldots, a_k$ are available. On repetition of the experiment they would, according to our theory, vary at random in concordance with the distribution 5.3, which depends on ξ. Also a random variable is $a_.$, the distribution of which 5.12 depends on ξ, and therefore it can be used for the estimation. But what 5.17 tells is that *the constellation of 0's and 1's* producing $a_.$, which also varies at random, *has a distribution that does not depend on* ξ. From the fundamental principle it then follows that once $a_.$ has been recorded, any extra information about *which of the items* were answered correctly is, according to our model, *useless as a source of inference about* ξ (but not for other purposes, as will presently be seen).

*In algebra they are known as elementary symmetric functions of $\epsilon_1, \ldots, \epsilon_k$.

The capital discovery that such situations exist was made by R. A. Fisher in 1922, and following his terminology we shall call $a_.$ *a sufficient statistic*—or *estimator*—for the parameter in question.

In the present situation, however, the sufficiency of $a_.$ needs a qualification as being *relative*, since it rests upon the condition that the ϵ_i's are known. As long as such knowledge is not available, the sufficiency as such is not very helpful, but the important point of 5.17 then is that it depends solely upon the ϵ's, not on ξ.

From 5.17 we can therefore proceed, as we did from 4.8, to consider a collection of subjects that all happen to have $a_. = r$. Specifying by $a_{\nu i}$ the a_i of subject ν and denoting by $(a_{\nu i})$, given ν, the set of responses $a_{\nu 1}, \ldots, a_{\nu k}$, that is,

(5.18) $$(a_{\nu i}) = (a_{\nu 1}, \ldots, a_{\nu k}),$$

we can rewrite 5.17 in the form

(5.19) $$p\{(a_{\nu i}) \mid a_{\nu .} = r\} = \frac{\epsilon_1^{a_{\nu 1}} \ldots \epsilon_k^{a_{\nu k}}}{\gamma_r}, \quad \nu = 1, \ldots, n.$$

The responses of the n persons being independent, their joint probability is obtained by multiplying the n probabilities of 5.19. Denoting for short the whole set of $n \times k$ responses by $((a_{\nu i}))$—the double parentheses indicating variation over both ν and i—we get

(5.20) $$p\{((a_{\nu i})) \mid (a_{\nu .} = r)\} = \frac{\epsilon_1^{a_{.1}} \ldots \epsilon_k^{a_{.k}}}{\gamma_r^n},$$

where

(5.21) $$a_{.i} = \sum_{\nu=1}^{n} a_{\nu i}.$$

Statement 5.20 implies that, as a consequence of the model, we have to deal with the total number of correct answers to each item for the n persons in question.

6. Separation of parameters

Let us finally consider the responses of n individuals with the parameters $\xi_1, \ldots, \xi_n$ to k items with the parameters $\epsilon_1, \ldots, \epsilon_k$. With the notation $a_{\nu i}$ introduced in the last section, the model 5.2 now takes the form

$$(6.1) \qquad p\{a_{\nu i} \mid \xi_\nu, \epsilon_i\} = \frac{(\xi_\nu \epsilon_i)^{a_{\nu i}}}{1+\xi_\nu \epsilon_i},$$

and on the assumption of stochastic independence of all of the responses $a_{\nu i}$, $\nu = 1, \ldots n$, $i = 1, \ldots k$, the joint probability of the whole set $((a_{\nu i}))$ of them becomes

$$p\{((a_{\nu i})) \mid (\xi_\nu),(\epsilon_i)\} = \prod_{\nu=1}^{n} \prod_{i=1}^{k} p\{a_{\nu i} \mid \xi_\nu, \epsilon_i\}$$

$$(6.2) \qquad = \frac{\prod_{\nu=1}^{n} \prod_{i=1}^{k} (\xi_\nu \epsilon_i)^{a_{\nu i}}}{\prod_{\nu=1}^{n} \prod_{i=1}^{k} (1+\xi_\nu \epsilon_i)}.$$

In the numerator we notice that the parameter ξ_ν occurs in k places, each time raised to a power $a_{\nu i}$, which all together makes $\xi_\nu^{a_{\nu.}}$, and that the parameter ϵ_i occurs in n places, each time raised to a power $a_{\nu i}$, adding up to a total power of $a_{.i}$. If furthermore the denominator is denoted by

$$(6.3) \qquad \gamma((\xi_\nu),(\epsilon_i)) = \prod_{\nu=1}^{n} \prod_{i=1}^{k} (1+\xi_\nu \epsilon_i),$$

we can simplify 6.2 to

$$(6.4) \qquad p\{((a_{\nu i})) \mid (\xi_\nu),(\epsilon_i)\} = \frac{\prod_{\nu=1}^{n} \xi_\nu^{a_{\nu.}} \cdot \prod_{i=1}^{k} \epsilon_i^{a_{.i}}}{\gamma((\xi_\nu),(\epsilon_i))}.$$

This formula is the generalization of 5.3 to n persons, but in consequence of 6.4 we now have to derive the probability that $a_{1.}, \ldots, a_{n.}$ and $a_{.1}, \ldots, a_{.k}$ take on two specified sets of values: $r_1, \ldots, r_n$ and $s_1, \ldots, s_k$.

In analogy to Section 5—in particular the logical chain of 5.11 through 5.13—we should find all possible ways of building up zero/one matrices $((a_{\nu i}))$ that have the same row totals $a_{\nu.} = r_\nu$, $\nu = 1, \ldots, n$ and column totals $a_{.i} = s_i$, $i = 1, \ldots, k$, state the probability of each realization, and add up all such probabilities to a total joint probability of the two sets of totals considered. However, this procedure is greatly simplified by the fact that all the probabilities to be added are equal, namely—according to 6.4—

$$(6.5) \qquad \frac{\prod_{\nu=1}^{n} \xi_\nu^{r_\nu} \cdot \prod_{i=1}^{k} \epsilon_i^{s_i}}{\gamma((\xi_\nu),(\epsilon_i))}.$$

Thus we have only to count the number of different ways in which it is algebraically possible to build up a zero/one matrix with the row totals of $r_\nu, \nu = 1, \ldots, n$ and the column totals of s_i, $i = 1, \ldots, k$.

Determining this number is a combinatorial problem that appears to be rather difficult, but at present we need nothing more than a notation. For this number we write

$$(6.6) \qquad \begin{bmatrix} r_1, \ldots, r_n \\ s_1, \ldots, s_k \end{bmatrix} = \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix},$$

and then we have

$$(6.7) \qquad p\{(a_{\nu.} = r_\nu),(a_{.i} = s_i) \mid (\xi_\nu),(\epsilon_i)\} = \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix} \frac{\prod_{\nu=1}^{n} \xi_\nu^{r_\nu} \prod_{i=1}^{k} \epsilon_i^{s_i}}{\gamma((\xi_\nu),(\epsilon_i))}.$$

This joint probability distribution of the row totals $a_{\nu.}$ and the column totals $a_{.i}$ contains just as many parameters as observables, and the latter would therefore seem suitable for estimation purposes. How true this is becomes clear when we divide 6.7 into 6.4—or 6.5—to obtain the probability of the whole set of observations, *on the condition that the totals of rows and columns are given.* In fact, all parametric terms cancel, and we are left with a conditional probability

$$(6.8) \qquad p\{((a_{\nu i})) \mid (a_{\nu.} = r_\nu),(a_{.i} = s_i)\} = \frac{1}{\begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix}}$$

that is independent of all of the parameters.

Therefore, once the totals have been recorded, any further statement as regards *which of the items* were answered correctly by *which persons* is, according to our model, *useless as a source of information about the parameters.* (Another use that can be made of the $a_{\nu i}$'s will emerge at a later stage of our discussion.) Thus the row totals and the column totals are *not only suitable* for estimating the parameters; they *imply every possible statement about the parameters that can be made on the basis of the ob-*

servations $((a_{\nu i}))$. Accordingly we will, in continuation of the terminology introduced in Section 5, characterize the row totals $a_{\nu.}$, $\nu = 1, \ldots, n$ and the column totals $a_{.i}$, $i = 1, \ldots, k$ as a *set of sufficient estimators for the parameters* $\xi_1, \ldots, \xi_n$ and $\epsilon_1, \ldots, \epsilon_k$.

As 6.7 contains both sets of parameters, a direct utilization of this formula would apparently lead to a simultaneous estimation of both sets. However, in view of previous results (see the comments following 5.17) it would seem appropriate to ask whether it is possible—also in this general case—to estimate the item parameters independently of the person parameters and, if so, vice versa as well.

In order to approach this problem we will first derive the distribution of the row totals, appearing as exponents of the ξ's, irrespective of the values of the column totals, by summing 6.7 over all possible combinations of $s_1, \ldots, s_k$. During this summation the denominator $\gamma((\xi_\nu), (\epsilon_i))$ remains constant, as do the terms $\xi_\nu^{r_\nu}$, $\nu = 1, \ldots, n$ in the numerator. Thus, on introducing the notation

(6.9) $$\gamma_{(r_\nu)}((\epsilon_i)) = \sum_{(s_i)} \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix} \epsilon_1^{s_1} \ldots \epsilon_k^{s_k}$$

we obtain

(6.10) $$p\{(a_{\nu.} = r_\nu) \mid (\xi_\nu), (\epsilon_i)\} = \frac{\gamma_{(r_\nu)}((\epsilon_i)) \cdot \prod_{\nu=1}^{n} \xi_\nu^{r_\nu}}{\gamma((\xi_\nu), (\epsilon_i))},$$

from which it is seen that the ξ_ν's might be estimated from the row totals *if the* ϵ_i*'s*—and therefore also the polynomials 6.9—*were known.*

Similarly, we can sum 6.7 over all possible combinations of $r_1, \ldots, r_n$, keeping $s_1, \ldots, s_k$ fixed. Substituting, in 6.9, $\xi_1, \ldots, \xi_n$ for $\epsilon_1, \ldots, \epsilon_k$ and in consequence interchanging the r's and the s's, we get

(6.11) $$\gamma_{(s_i)}((\xi_\nu)) = \sum_{(r_\nu)} \begin{bmatrix} (s_i) \\ (r_\nu) \end{bmatrix} \xi_1^{r_1} \ldots \xi_n^{r_n},$$

where, by the way,

(6.12) . $$\begin{bmatrix} (s_i) \\ (r_\nu) \end{bmatrix} = \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix}.$$

With this notation the summation yields on analogy to 6.10:

(6.13) $$p\{(a_{.i} = s_i) \mid (\xi_\nu), (\epsilon_i)\} = \frac{\gamma_{(s_i)}((\xi_\nu)) \cdot \prod_{i=1}^{k} \epsilon_i^{s_i}}{\gamma((\xi_\nu), (\epsilon_i))},$$

and accordingly the ϵ_i's might be estimated from the column totals *provided the* ξ_ν*'s were known.*

Thus we might estimate the ξ's if the ϵ's were known, and the ϵ's if the ξ's were known! And both estimations would even be relatively sufficient. In fact, on dividing 6.10 into 6.7 to obtain the conditional probability of $a_{.i}$ for given $a_{\nu.}$, we get

(6.14) $$p\{(a_{.i} = s_i) \mid (a_{\nu.} = r_\nu), (\xi_\nu), (\epsilon_i)\} = \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix} \frac{\prod_{i=1}^{k} \epsilon_i^{a_{.i}}}{\gamma_{(r_\nu)}((\epsilon_i))},$$

which is independent of the parameters ξ_ν to be estimated. Similarly, the division of 6.13 into 6.10 gives

(6.15) $$p\{(a_{\nu.} = r_\nu) \mid (a_{.i} = s_i), (\xi_\nu), (\epsilon_i)\} = \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix} \frac{\prod_{\nu=1}^{n} \xi_\nu^{r_\nu}}{\gamma_{(s_i)}((\xi_\nu))},$$

which is independent of the ϵ's. But, of course, as long as neither set of parameters is known, these possibilities are of no avail.

It is one of the characteristic features of the model under consideration that this vicious circle can be broken, the instrument being a reinterpretation of the formulas 6.14 and 6.15. In fact, as 6.14 depends on the ϵ's but not on the ξ's, this formula gives the opportunity of estimating the ϵ's without dealing with the ξ's. Thus the objections to both 6.7 and 6.13 have been eliminated. The unknown ξ's in these expressions have been replaced with observable quantities: the individual totals $a_{\nu.}$. Similarly, in 6.15 the ϵ's of 6.7 and 6.10 have been replaced with the item totals $a_{.i}$, in consequence of which we can estimate the ξ's without knowing or simultaneously estimating the ϵ's. Thus the estimation of the two sets of parameters can be separated from each other.

In this connection we can return to 6.8, noting that this formula is a consequence of the model structure—3.3 and the stochastic independence—irrespective of the values of the parameters of which the right-hand term is independent. Therefore, if from a given matrix $((a_{\nu i}))$ we construct a quantity that would be useful for disclosing a particular type of departure from the model, then its sampling distribution as conditioned by the marginals $a_{\nu.}$ and

$a_{.i}$ will be independent of all of the parameters. Thus the testing of the model can be separated from dealing with the parameters.

We will not consider here the question of how to perform such testing in practice and also that of turning the observed row and column totals into adequate estimates of the ξ's and the ϵ's.

In Chapter 6 of the earlier study (3) these questions were dealt with by simple methods which were taken to be acceptable approximations. In the case of subtest N the observations passed the test for the model satisfactorily, but the model failed completely in the case of subtest F (geometrical shapes). In the latter subtest the time allowance, for some technical reasons, had been cut below the optimal limit, but a reanalysis of the data (not reported here) has shown that when allowance is made for the working speed of each subject, the data fit the model just as well as for the numerical sequences.

However, from a theoretical point of view the method used to test the model was unsatisfactory (see Rasch [3], Chapter 10, in particular pp. 181–182). By now we are in the process of working out better methods, and therefore for the time being we shall leave the documentation of the applicability of the model with simply a reference to the earlier work.

7. Specific objectivity

As regards the basic formulas 6.14 and 6.15, we have already noted that when they are applied to the total set of data they enable us to separate the estimation of one set of parameters from that of the other. However, formula 6.15 can also be applied to any subgroup of the total collection of subjects having been exposed to the k stimuli. Thus the parameters of the subjects in the subgroup can be evaluated without regard to the parameters of the other subjects.

In particular, the parameters of any two subjects can be compared on their own virtues alone, quite irrespective of the group or population to which for some reason they may be referred. Thus, as indicated in Section 1, the new approach, when applicable, does rule out populations from the comparison of individuals.

Similarly, formula 6.14 can be applied to any subset of the k stimuli, and accordingly their parameters can be evaluated without regard to the parameters of the other stimuli. In particular, the parameters of any two stimuli can be compared separately.

With these additional consequences, the principle of separability leads to a singular objectivity in statements about both parameters and model structure. In fact, *the comparison of any two subjects* can be carried out in such a way that *no other parameters are involved than those of the two subjects*—neither the parameter of any other subject nor any of the stimulus

parameters. Similarly, *any two stimuli can be compared independently of all other parameters than those of the two stimuli*, the parameters of all other stimuli as well as the parameters of the subjects having been replaced with observable numbers.

It is suggested that comparisons carried out under such circumstances be designated as "specifically objective." The same term would seem appropriate for statements about the model structure that are independent of all the parameters specified by the model, their unknown values being, in fact, irrelevant for the structure of the model.

Of course, specific objectivity is no guarantee against the "subjectivity" of the statistician when he chooses his fiducial limits or when he judges which kind of deviations from the model he will look for. Neither does it save him from the risk of being offered data marred by the subjective attitude of the psychologist during his observations. Altogether, when introducing the concept of specific objectivity, I am not entering upon a general philosophical debate on the meaning and the use of objectivity at large. At present the term is strictly limited to observational situations that can be covered by the stimulus-subject-response scheme, to be described in terms of a parametric model that specifies parameters for stimuli and for subjects.

What has been demonstrated in detail in the case of two response categories is *that the specified objectivity in all three directions can be attained insofar as the type of model defined herein holds*. Recently it has been shown that except for unimportant mathematical restrictions, the inverse statement is also true: if only two responses are available, the observations must conform to the simple model 3.3 if it is to be possible to maintain specific objectivity in statements about subjects, stimuli, and model.

8. Fields of application

The problems we have been dealing with in the present paper were formulated within a narrow field, psychological test theory. However, with the discovery of specific objectivity we have arrived at concepts of such generality that the original limitation is no longer justified. Extensions into other fields of psychology, such as psychophysical threshold experiments and experiments on perception of values, offer themselves, but the stimulus-subject-response framework is by no means limited to psychology. Thus in a recent publication (1) a Poisson model was employed in an investigation of infant mortality in Denmark in the period 1931–60. In each year the number of infant deaths (of all causes or of a particular cause) out of the number of boys and of girls, born in or out of wedlock, was recorded. In this case the years served as subjects, the combination of sex and legitimacy of the children as the stimulus, and the number of infant deaths out of the number of children born as the responses.

From economics we can take household budgets as an example. The families serve as subjects, income and expenditures—classified into a few types—as stimuli, and the amount earned and the amount spent as the responses.

These examples may indicate that the framework covers a rather large field within the social sciences. Delineating the area in which the models described here apply is a huge problem, the inquiry into which has barely started.

But already the two intelligence tests mentioned in Section 2 and discussed at the end of Section 6 are instructive as regards the sort of difficulties we should be prepared to meet. For one of them, the numerical sequences, the earlier analysis (3, Chapter 6) showed a perfectly satisfactory fit of the observations to the model—that is, in this case specific objectivity can be obtained on the basis of the response pattern for each subject. For the other test, the geometrical shapes, the analysis most unambiguously showed that the separability did not hold.

Neither did it hold for a different intelligence test, which was of the omnibus type, containing items that presumably called upon very different intelligence functions. In this case, therefore, the data could not be expected to allow for a description comprising only one parameter for each subject. The items in the numerical sequences were quite uniform in that they required the testee to recognize a logical structure in a sequence of numbers. According to the analysis, the items were sufficiently uniform—although of very different levels of difficulty—to allow for a description of the data by one parameter for each subject as well as for each item. The items of the geometric figure test were constructed just as uniformly as the numerical sequences, and therefore it was somewhat of a surprise that the results turned out quite adversely.

To this material I could add observations on two other tests, constructed with equal care. One was a translation of the idea of Raven's matrix test into letter combinations, at the same time substituting the multiple choice by a construction, on the part of the testee, of the answer. For this test the results were just as good as for the numerical sequences. The other test consisted of a set of verbal analogies where the number of answers offered was practically infinite, with the effect that the multiple choice was in fact eliminated. Here the results of the testing were just as disappointing as for the figure test.

This contrast, however, led to the solution of the mystery. The difference between the two pairs of tests was due not to construction principles but to the administration of the tests. For all four tests the adequate time allowance was determined by means of special experiments. On applying them to random samples of 200 subjects, it turned out that the number of correct answers formed a convenient distribution for the letter matrix test and for the numerical sequences, but verbal analogies and the figure test were too

easy and the distributions showed an undesirable accumulation of many correct answers.

This happened in 1953, when only the barest scraps of the theory had been developed, and yielding to considerable time pressure the test constructor, consulting me on the statistical part of the problem, severely cut down the time allowances so as to move the distributions to the middle of the range. While succeeding in that, we spoiled the test, turning it into a mixture of a test for ability and a test for speed.

More recently, however, I have had the opportunity to reanalyze both sets of data, grouping the subjects primarily according to their working speed, as given by the number of items done, and applying to each group the technique of Chapter 6 of the earlier report (2). The result was startling: Within each speed group I found confirmation of the theory, and the relative difficulties of the items were independent of the working speed. Altogether, with speed as ancillary information, specific objectivity can be attained in regard to the properties which the tests really aimed at measuring.

Inverting the final statement, we get the moral of this story: *Observations may easily be made in such a way that specific objectivity, otherwise available, gets lost.* For instance, this can easily happen when qualitative observations with, say, five categories of responses for convenience are condensed into three categories. If the basic model holds for the five categories, it is mathematically almost impossible for the three-category model also to hold. Thus the grouping, tempting as it may be, will usually tend to slur the specific objectivity.

In concluding, therefore, I must point out that the problem of the relation of data to models is not only one of trying to fit data to an adequately chosen model from our inventory to see whether it works; it is also *how to make observations in such a way that specific objectivity obtains.*

References

1. MATTHIESSEN, P. C. *Infant Mortality in Denmark, 1931–60.* Copenhagen: Statistical Department, 1965.
2. RASCH, G. "On General Laws and the Meaning of Measurement in Psychology," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* Vol. IV, 321-34, Berkeley: Univ. of California Press, 1961.
3. ———. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research, 1960.