DOCUMENT RESUME

ED 053 162 TM 000 679

|  |  |
|---|---|
| AUTHOR | Urry, Vern W. |
| TITLE | Individualized Testing by Bayesian Estimation. |
| INSTITUTION | Washington Univ., Seattle. Bureau of Testing. |
| PUB DATE | Apr 71 |
| NOTE | 31p. |
| EDRS PRICE | EDRS Price MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Achievement Tests, *Bayesian Statistics, *Computer Programs, *Educational Testing, Factor Analysis, Factor Structure, Guessing (Tests), *Mathematical Models, Measurement Techniques, Research Methodology, Simulation, Statistical Analysis, Statistical Data, *Test Construction, Testing, Validity |
| IDENTIFIERS | *Tailored Testing |

ABSTRACT

Bayesian estimation procedures are summarized and numerically illustrated by means of simulation methods. Procedures of data generation for simulation purposes are also delineated and computationally demonstrated. The logistic model basic to the Bayesian estimation procedures is shown to be explicit with respect to the probability distribution from which one is sampling. This feature allows for an assessment or evaluation of its capabilities without empirical data. The fit of the model to empirical data is discussed as an issue independent of considerations as to model capabilities. Three item banks are used to simulate Bayesian estimation procedures. Two are idealized--though reasonably possible--examples; whereas, the third consists of items specified according to other parameter estimates reported. Results indicate that with test validity held constant, Bayesian tailored testing of the Verbal Scholastic Aptitude Test (VSAT) could result in a savings of 65% of testing time for the average examinee. However, more savings in testing time is seen as possible through the use of item banks developed specifically for the purpose of tailored testing. The present investigation did not utilize prior information. Further assessments of model capabilities should explore such usage. While present results appear favorable, the full potentialities of the model have yet to be assessed. (Author/AE)

*Bureau of Testing*
*University of Washington—Seattle*

Bureau of Testing

University of Washington

April 1971

# Individualized Testing by Bayesian Estimation

Vern W. Urry

Bayesian estimation procedures derived by Owen (1969) were summarized and numerically illustrated by means of simulation methods. Procedures of data generation for simulation purposes were also delineated and computationally demonstrated.

The logistic model basic to the Bayesian estimation procedures was shown to be explicit with respect to the probability distribution from which one is sampling. This feature of the model allows for an assessment or evaluation of its capabilities <u>sans</u> empirical data. The fit of the model to empirical data was discussed as an issue independent of considerations as to model capabilities.

Three item banks were used to simulate Bayesian estimation procedures. Two of the banks were idealized--though reasonably possible--examples; whereas, the third consisted of items specified according to parameter estimates reported by Lord (1968) for the VSAT.

With test validity held constant, Bayesian tailored testing of the VSAT could result in a savings of 65% of testing time for the average examinee. However, more savings in testing time was viewed as possible through the use of item banks developed specifically for the purpose of tailored testing.

The present investigation did not utilize prior information. Further assessments of model capabilities should explore such usage. While present results appear favorable, the full potentialities of the model have yet to be assessed.

Bureau of Testing Project: 0171-177

## Individualized Testing by Bayesian Estimation

Owen (1969) has derived Bayesian procedures for the tailoring of tests for the cases where chance success on the items is and is not effective. Both cases will be discussed in the current report along with illustrative data. Computer programs which simulate the process are described and included in the appendix for the separate cases. The programs can be modified for "live" tailored testing applications.

Under both cases, the procedures: (1) identify the most appropriate item for presentation; (2) score the response to that item or, synonomously, (re-) estimate the ability parameter for the individual; and (3) calculate the standard error of the new estimate of ability. The process can be repeated until all or a specified number of items have been used or an allowable value of the standard error of estimate has been attained.

In the following, we assume that the item parameters are either known or have been previously estimated. By way of review, the item parameters of the logistic model are item discriminatory power ($a_i$); item difficulty ($b_i$); and probability of chance success on the item ($c_i$). Methods of estimating the item parameters have been discussed by Birnbaum (1968).

### Method

In both cases, one calculates $\hat{\theta}_{(p)}$, the estimate of ability, and $\hat{\sigma}^2_{(p)}$, the variance of the estimate, sequentially. The subscript p indexes the number of items that have been presented to an individual during an evaulation sequence. For example, if one lacks prior information on the individual being examined, $\hat{\theta}_{(0)}$ and $\hat{\sigma}^2_{(0)}$ would be set at values of 0 and 1, respectively. Initial values of this nature have the basic rationale that in the absence of individual information, the mean, 0, is the most probable estimate of ability, $\hat{\theta}_{(0)}$, while the standard error, $\hat{\sigma}_{(0)}$, coincides with the standard deviation. More concisely expressed, the prior distribution of ability, $\theta$, is assumed to be $N(0,1)$.

Case I: Chance success on the items is not effective

In order to determine the item most appropriate for immediate presentation, we calculate $\alpha_i$ for all (unused) items. The formula is

$$(1) \qquad \alpha_i = [a_i^{-2} + \hat{\sigma}^2_{(p)}] \exp(2D_i^2)[1 - (\text{erf } D_i)^2]$$

where

$$(2) \qquad D_i = (b_i - \hat{\theta}_{(p)}) \{2[a_i^{-2} + \hat{\sigma}^2_{(p)}]\}^{-\frac{1}{2}}$$

and

$$(3) \qquad \text{erf } D_i = \frac{2}{\sqrt{\pi}} \int_0^{D_i} \exp(-t^2)\, dt \quad .$$

More familiarly, we have

$$(4) \qquad \text{erf } D_i = 2\Phi(\sqrt{2}D_i) -1$$

where

$$(5) \qquad \Phi(\sqrt{2}D_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{2}D_i} \exp\left[-\frac{t^2}{2}\right] dt$$

is the well tabled normal probability function. For future reference and convenience, we will designate the following:

$$(6) \qquad s_i = \frac{\hat{\sigma}^2_{(p)}}{\sqrt{a_i^{-2} + \hat{\sigma}^2_{(p)}}} \quad .$$

(7) $$t_i = [1 + a_i^{-2}\,\hat{\sigma}_{(p)}^{-2}]^{-1}$$

(8) $$u_i = (1 - \mathrm{erf}\ D_i)$$

(9) $$v_i = (1 + \mathrm{erf}\ D_i)$$

Now the subscript, i, of the smallest $\alpha_i$ given by equations (1) identifies the optimal item for presentation. Upon presentation of the $i^{\underline{th}}$ item, one of two outcomes is possible. The individual will get the item right, or he will get the item wrong. Given that he responded correctly, his new ability estimate is given by

(10) $$\hat{\theta}_{(p+1)}(\text{right}) = \hat{\theta}_{(p)} + \sqrt{\frac{2}{\pi}}\,\exp(-D_i^2)\ s_i u_i^{-1} \quad .$$

The standard error of the estimate, conditional upon a correct response, is given by the square root of

(11) $$\hat{\sigma}^2_{(p+1)}(\text{right}) = \hat{\sigma}^2_{(p)}\{1 - \frac{2}{\sqrt{\pi}}\,t_i[\exp\ (D_i^2)u_i]^{-2} \cdot [\frac{1}{\sqrt{\pi}} - D_i\ \exp\ (D_i^2)u_i]\} \quad .$$

Should the individual miss the item, his new ability estimate is given by

(12) $$\hat{\theta}_{(p+1)}(\text{wrong}) = \hat{\theta}_{(p)} - \sqrt{\frac{2}{\pi}}\ \exp(-D_i^2)s_i v_i^{-1} \quad .$$

The standard error of the new estimate of ability, conditional upon an incorrect response, is given by the square root of

(13) $$\hat{\sigma}^2_{(p+1)}(\text{wrong}) = \hat{\sigma}^2_{(p)}\{1 - \frac{2}{\sqrt{\pi}}\,t_i\ [\exp\ (D_i^2)v_i]^{-2} \cdot [\frac{1}{\sqrt{\pi}} + D_i\ \exp(D_i^2)v_i]\} \quad .$$

At this point, the number of item presentations (p) is updated by one since the sequential use of equations (1) and equations (10) and (11) or equations (12) and (13) defines an item presentation cycle. As a consequence, the current $\hat{\theta}_{(p+1)}$ and $\hat{\sigma}^2_{(p+1)}$ become the $\hat{\theta}_{(p)}$ and $\hat{\sigma}^2_{(p)}$ when a new item presentation cycle is initiated. When equations (1) are recalculated for the (n - p) unused items, the $i^{\underline{th}}$ item is again identified by the smallest $\alpha_i$ . Again, depending on the propriety of the individual's response either equations (10) and (11) or equations (12) and (13) are used. The cycles may be repeated until a termination criterion has been attained.

Case II: Chance success on the items is effective

Equations analogous or even identical to those for Case I exist for the present case; however, due to the effectiveness of guessing, some equations increase in complexity. Let us designate

$$(14) \qquad w_i = [c_i + \frac{(1 - c_i)u_i}{2}]$$

for further refrrence. Now one begins by calculating

$$(15) \qquad \beta_i = [(1 - c_i)t_i]^{-1}w_i\ (1 - \frac{u_i}{2})\exp\ (2D_i^2)[1 + c_i(1 - \frac{u_i}{2})]^{-1} \quad ,$$

for all (unused) items where equations (2) through (9) still obtain. Again the subscript, i, of the smallest $\beta_i$ identifies the optimal item for immediate presentation.

One of two outcomes will occur when the $i^{\underline{th}}$ item is presented. If the individual responds correctly, his new estimate of ability is given by

$$(16) \qquad \hat{\theta}_{(p+1)}(\text{right}) = \hat{\theta}_{(p)} + \frac{(1 - c_i)}{\sqrt{2\pi}}\ \exp(-D_i^2)s_i w_i^{-1} \quad .$$

The standard error of the new estimate, conditional upon the appropriate response, is given by the square root of

$$(17) \qquad \sigma^2_{(p+1)}(\text{right}) = \sigma^2_{(p)} \left[ 1 - \frac{(1 - c_i)}{\sqrt{\pi}} (w_i u_i)^{-1} t_i \exp(-2D_i^2) \cdot \{ [ \frac{1}{\sqrt{\pi}} - D_i \exp(D_i^2) u_i ] - \frac{c_i}{\sqrt{\pi}} w_i^{-1} \} \right] \quad .$$

Should the individual respond incorrectly to the $i^{\underline{th}}$ item, equations (12) and (13) are still appropriate; however, for convenience the equations are repeated. His new estimate of ability is provided by

$$(18) \qquad \hat{\theta}_{(p+1)}(\text{wrong}) = \hat{\theta}_{(p)} - \sqrt{\frac{2}{\pi}} \exp(-D_i^2) s_i v_i^{-1} \quad .$$

The standard error of the new ability estimate, given a wrong answer, is obtained from the square root of

$$(19) \qquad \hat{\sigma}^2_{(p+1)}(\text{wrong}) = \hat{\sigma}^2_{(p)} \{ 1 - \frac{2}{\sqrt{\pi}} t_i [\exp (D_i^2) v_i]^{-2} \cdot [ \frac{1}{\sqrt{\pi}} + D_i \exp(D_i^2) v_i ] \} \quad .$$

The sequential process of item presentation cycles delineated above is also appropriate here. The optimal item for immediate presentation from among the (n - p) unused items is determined for each cycle by the subscript on the smallest $\beta_i$ as given by equations (15). The $i^{\underline{th}}$ item is responded to by the individual and the nature of his response determines whether equations (16) and (17) or (18) and (19) are to be used in estimating ability and the variance of the estimate of ability. As indicated above, termination criteria may be specified on the basis of a maximum value for p and a maximum allowable value for $\hat{\sigma}_{(p)}$ .

Under both cases, the estimates $\hat{\theta}_{(p+1)}$ and $\hat{\sigma}^2_{(p+1)}$ are the first and second moments of the posterior distribution. When these same estimates are used to determine the next item to be presented their status changes to $\hat{\theta}_{(p)}$ and $\hat{\sigma}^2_{(p)}$, the first and second moments of the prior distribution. In Bayesian estimation procedures, the posterior distribution becomes the prior distribution when a new item presentation cycle is initiated.

## Generation of response vectors for simulated individualized testing

Given any set of n items having known item parameters, the probability distribution, conditional upon ability, $\theta$, can be determined for all possible response patterns or vectors. What this means is that if one has a random sample of values of $\theta$, one can then sample response vectors and score these by the procedural manner he chooses. Later we will discuss how one can evaluate the given procedure. In this instance, we choose to evaluate Bayesian estimation procedures, but the simulation technique has wider applicability. A case in point would be an evaluation of flexilevel testing (Lord, 1971).

For purposes of concrete illustration, we will take a 4-item example. Given four items, there are $2^4$ or 16 possible patterns or response vectors, $v_k$. These are:

$$v_1 = [0\ 0\ 0\ 0]$$
$$v_2 = [0\ 0\ 0\ 1]$$
$$v_3 = [0\ 0\ 1\ 0]$$
$$v_4 = [0\ 0\ 1\ 1]$$
$$v_5 = [0\ 1\ 0\ 0]$$
$$v_6 = [0\ 1\ 0\ 1]$$
$$v_7 = [0\ 1\ 1\ 0]$$
$$v_8 = [0\ 1\ 1\ 1]$$
$$v_9 = [1\ 0\ 0\ 0]$$
$$v_{10} = [1\ 0\ 0\ 1]$$

$$v_{11} = [1\ 0\ 1\ 0]$$
$$v_{12} = [1\ 0\ 1\ 1]$$
$$v_{13} = [1\ 1\ 0\ 0]$$
$$v_{14} = [1\ 1\ 0\ 1]$$
$$v_{15} = [1\ 1\ 1\ 0]$$
$$v_{16} = [1\ 1\ 1\ 1]$$

Now response vector $v_{11}$ indicates that items 1, 2, 3, and 4 were responded to correctly, incorrectly, correctly and incorrectly, respectively. If, further, the item parameters for the items of illustration are:

| | $a_i$ | $b_i$ | $c_i$ |
|---|---|---|---|
| Item 1 | 1.6 | 1.1 | .06 |
| Item 2 | 2.0 | 1.1 | .05 |
| Item 3 | 1.2 | 1.1 | .05 |
| Item 4 | 1.0 | 1.1 | .13 |

and, say, the value of $\theta$ for which one is evaluating the probability distribution of the $v_k$ is 1.0, we would proceed as follows. The probability of a correct response to an item, given $\theta$, is provided by the model as

$$(20) \qquad P_i(\theta) = c_i + (1 - c_i)\frac{1}{1 + \exp[-Da_i(\theta - b_i)]}$$

where D is the constant 1.7. The probability of missing an item is merely

$$(21) \qquad Q_i(\theta) = 1 - P_i(\theta) \quad .$$

As a consequence, the following probabilities obtain:

$$P_1(\theta = 1.0) = .47$$
$$Q_1(\theta = 1.0) = .53$$

$$P_2(\theta = 1.0) = .45$$
$$Q_2(\theta = 1.0) = .55$$

$$P_3(\theta = 1.0) = .48$$
$$Q_3(\theta = 1.0) = .52$$

$$P_4(\theta = 1.0) = .53$$
$$Q_4(\theta = 1.0) = .47 \quad .$$

Since ability is fixed, the probabilities are assumed independent across items. This assumption is more familiarly known as that of local independence. It merely states that if several variables covary with one another due to a distinct variable, holding the latter constant results in independence among the several. We may now compute the joint probability of the independent events indicated by each $v_k$ for $\theta$ equal to 1.0. This is merely the product of the probabilities of the events recorded by the zeros and ones in each $v_k$ . For convenience, the probabilities conditional upon $\theta$ are calculated in Table 1. For example, response vector $v_{11}$ recorded that: Item 1 was correct with probability equal to $P_1(\theta = 1.0)$, or .47; Item 2 was incorrect with probability equal to $Q_2(\theta = 1.0)$, or .55; Item 3 was correct with probability equal to $P_3(\theta = 1.0)$, or .48; and Item 4 was incorrect with probability equal to $Q_4(\theta = 1.0)$ or .47. The probability of the joint events conditional upon $\theta = 1.0$ is, then, (.47)(.55)(.48)(.47) or .0583. In Table 1, the cumulative conditional probabilities are also given. Handily, as well as properly, these sum to unity. As a consequence, one can obtain a randomly selected response vector for a given value of $\theta$ by obtaining a random number from a distribution which is uniform on the interval from zero to unity and comparing this to the attendant probability intervals. For the sake of clarification, let us say that a random number thusly selected was .6254. Since the value occurs in the probability interval of .5933 to .6644 corresponding to response vector $v_{10}$ , the said vector would have been randomly selected in proportion to its probable occurrence given the stated conditions. Clearly, the nature of the sampling remains unchanged even with an arbitrary ordering of the response vectors.

Table 1

Conditional Probabilities, Cumulative Conditional Probabilities, and Probability Intervals for Possible Response Vectors

| Conditional Probabilities | Cumulative Conditional Probabilities | Probability Intervals |
|---|---|---|
| $\text{Prob}(v_1 \mid \theta = 1.0) = (.53)(.55)(.52)(.47) = .0712$ | .0712 | .0000 to .0712 |
| $\text{Prob}(v_2 \mid \theta = 1.0) = (.53)(.55)(.52)(.53) = .0803$ | .1515 | .0713 to .1515 |
| $\text{Prob}(v_3 \mid \theta = 1.0) = (.53)(.55)(.48)(.47) = .0658$ | .2173 | .1516 to .2173 |
| $\text{Prob}(v_4 \mid \theta = 1.0) = (.53)(.55)(.48)(.53) = .0742$ | .2915 | .2174 to .2915 |
| $\text{Prob}(v_5 \mid \theta = 1.0) = (.53)(.45)(.52)(.47) = .0583$ | .3498 | .2916 to .3498 |
| $\text{Prob}(v_6 \mid \theta = 1.0) = (.53)(.45)(.52)(.53) = .0657$ | .4155 | .3499 to .4155 |
| $\text{Prob}(v_7 \mid \theta = 1.0) = (.53)(.45)(.48)(.47) = .0538$ | .4693 | .4156 to .4693 |
| $\text{Prob}(v_8 \mid \theta = 1.0) = (.53)(.45)(.48)(.53) = .0607$ | .5300 | .4694 to .5300 |
| $\text{Prob}(v_9 \mid \theta = 1.0) = (.47)(.55)(.52)(.47) = .0632$ | .5932 | .5301 to .5932 |
| $\text{Prob}(v_{10} \mid \theta = 1.0) = (.47)(.55)(.52)(.53) = .0712$ | .6644 | .5933 to .6644 |
| $\text{Prob}(v_{11} \mid \theta = 1.0) = (.47)(.55)(.48)(.47) = .0583$ | .7227 | .6645 to .7227 |
| $\text{Prob}(v_{12} \mid \theta = 1.0) = (.47)(.55)(.48)(.53) = .0658$ | .7885 | .7228 to .7885 |
| $\text{Prob}(v_{13} \mid \theta = 1.0) = (.47)(.45)(.52)(.47) = .0517$ | .8402 | .7886 to .8402 |
| $\text{Prob}(v_{14} \mid \theta = 1.0) = (.47)(.45)(.52)(.53) = .0583$ | .8985 | .8403 to .8985 |
| $\text{Prob}(v_{15} \mid \theta = 1.0) = (.47)(.45)(.48)(.47) = .0477$ | .9462 | .8986 to .9462 |
| $\text{Prob}(v_{16} \mid \theta = 1.0) = (.47)(.45)(.48)(.53) = .0538$ | 1.0000 | .9463 to 1.0000 |

The principles developed above generalize directly to situations where there are a large number of items; however, it is then usually more convenient to work with repeated samplings from subsets of items with $\theta$ fixed. Notice that with 100 items the calculation of $2^{100}$ conditional probabilities for fixed $\theta$ would present severe computational problems.

A computer program that accomplishes the sampling of response vectors has been presented elsewhere (Urry, 1970). More specifically, the program samples response vectors for a random sample from the assumed distribution of $\theta$, $N(0,1)$. Obviously, gaussian random numbers will fulfill the imposed sampling requirements with respect to underlying ability.

Notice that only a subset of items from any response vector of length $n$ are actually used by the Bayesian procedures to obtain an estimate of ability, $\hat{\theta}_{(p)}$. The item sequence, as noted above, is determined by prior information and/or responses as well as the item parameters. Given the model, the temporality of responses, as far as the n-length response vector is concerned, is inconsequential. In other words, one may (re-) estimate ability on any sequence or subset of items from the response vector as determined by a procedure while ignoring the available responses to the remaining items. The simulated or after-the-fact tailoring of empirically obtained response vectors is also possible.

Evaluation of the Bayesian Procedures

To evaluate the procedures, one merely correlates the $\hat{\theta}_{(p)}$ against the $\theta$ for the randomly sampled "cases" or simulated individuals for a particular termination criterion. Underlying ability, $\theta$, is the perfect criterion for the computation of this validity coefficient.

Regarding valid evaluation, the specifications for the item parameters of an item bank are critical. For example, high item discriminatory powers and a rectangular distribution of item difficulties have in other tailored testing contexts led to quite satisfactory validities. Here we will consider three item banks, two of which are idealized while the third is taken from an empirical source.

*Item Bank A* consisted of 100 items. The item discriminatory powers, $a_i$ , equalled 1.6 for all i, i = 1, 2, ... n or 100, while 20 items each had item difficulties, $b_i$, at one of five levels, i.e., -1.50, -.75, .00, .75, and 1.50, respectively. The probability of chance success on the items, $c_i$ , was .2 for all items.

*Item Bank B* consisted of 105 items. Again, the item discriminatory powers, $a_i$ , equalled 1.6 for all items, while 5 items each had item difficulties at one of twenty-one levels, i.e., -2.50, -2.25, -2.00 ... 2.50, respectively. The probability of chance success on the items, $c_i$ , was, again, .2 for all items.

The item parameters for *Item Bank C* were taken from estimates provided by Lord (1968) for the Verbal Scholastic Achievement Test (VSAT). The estimation sample was comprised of 2,862 cases. The interested reader will find the specifications for this item bank enumerated in that source.

With these specifications, simulated response vectors were generated, as previously outlined, for samples of 50 each for *Item Banks A* and *B*. For *Item Bank C*, 100 "cases" were generated. Termination criteria for *Item Banks A* and *B* were set at .32 and .25 as maximum allowable values for $\hat{\sigma}_{(p)}$ . Under the termination rules, p will vary across simulated examinees. For *Item Bank C*, the compound termination criterion was $p = 30$ or $\hat{\sigma}_{(p)}$ equal to or less than a maximum value of .25.

# Results

In order to illustrate what an individual evaluation sequence would look like, Table 2 is provided. The data are for a simulated "case" who was evaluated with Item Bank A. At the beginning of evaluation, nothing was assumed known about Examinee 1046; therefore, our prior information admits to this state of ignorance by setting $\hat{\theta}_{(0)}$ and $\hat{\sigma}_{(0)}$ to .00 and 1.00, respectively. In other words, the mean is, then, the most probable value or the best estimate of $\theta$ while the standard error of the estimate corresponds to the standard deviation of the prior distribution. Given this initial information, Item 3 from Item Bank A was found via equations (15) to be the most appropriate for the first presentation. The examinee answered Item 3 incorrectly, so that equations (18) and (19) were used to calculate the specific values of -.6766 and .7363, respectively for $\hat{\theta}_{(1)}$ and $\hat{\sigma}_{(1)}$. Using these estimates, equations (15) were evaluated for the (n - 1) or remaining 99 items. Item 2 from Item Bank A was, thereby, identified as the most appropriate item for the second presentation. The examinee answered Item 2 correctly; consequently, equations (16) and (17) were used to calculate $\hat{\theta}_{(2)}$ and $\hat{\sigma}_{(2)}$, or -.3904 and .6694, specifically and respectively. The cyclic processing continues, as indicated earlier, until a termination criterion is reached. At a termination criterion of $\hat{\sigma}_{(p)}$ less than or equal to .32, $\hat{\theta}_{(10)}$ or -.9937 was the Bayesian estimate of ability for Examinee 1046. Given the termination criterion of $\hat{\sigma}_{(p)}$ less than or equal to .25, $\hat{\theta}_{(14)}$ or -1.2104 was the Bayesian estimate of ability. Now the "true" value of $\theta$ was -1.2180 for the particular examinee. Notice that if we establish confidence intervals for $\theta$ with an approximate probability of .95, we have:

$$\text{Prob}\ [-.99 - (1.96)(.32) \le \theta \le -.99 + (1.96)(.32)] \cong .95$$

Table 2

An Example of an Individual Evaluation Sequence by Bayesian Estimation Procedures

Examinee Number 1046

Prior Information

$\hat{\theta}_{(0)} = .00$

$\hat{\sigma}_{(0)} = 1.00$

| Item | Item Bank | Item | Ability | Standard Error |
|---|---|---|---|---|
| Presentation (p) | Number | Response | Estimate [$\hat{\theta}_{(p)}$] | of Estimate [$\hat{\sigma}_{(p)}$] |
| 1 | 3 | 0 (wrong) | -.6766 | .7363 |
| 2 | 2 | 1 (right) | -.3904 | .6694 |
| 3 | 7 | 1 (right) | -.1902 | .6079 |
| 4 | 8 | 0 (wrong) | -.4718 | .5140 |
| 5 | 12 | 1 (right) | -.3330 | .4789 |
| 6 | 17 | 0 (wrong) | -.6717 | .4085 |
| 7 | 22 | 1 (right) | -.5597 | .3895 |
| 8 | 27 | 0 (wrong) | -.7594 | .3500 |
| 9 | 32 | 0 (wrong) | -.8944 | .3224 |
| 10 | 37 | 0 (wrong) | -.9937 | .3018 |
| 11 | 1 | 0 (wrong) | -1.1660 | .2791 |
| 12 | 6 | 1 (right) | -1.1230 | .2714 |
| 13 | 11 | 0 (wrong) | -1.2506 | .2549 |
| 14 | 16 | 1 (right) | -1.2104 | .2489 |

or

$$\text{Prob } [-1.62 \leq \theta \leq -.36] \cong .95$$

and

$$\text{Prob } [-1.21 - (1.96)(.25) \leq \theta \leq -1.21 + (1.96)(.25)] \cong .95$$

or

$$\text{Prob } [-1.70 \leq \theta \leq -.72] \cong .95$$

where $\theta$ is well within the indicated intervals. The probabilities are approximate since the $\hat{\sigma}_{(p)}$ are unbiased only when $p$ is large.

We now turn to results in samples where the individualized testing delineated above was applied to each "case." Validity coefficients for the three item banks are presented in Table 3. For example, using *Item Bank A* with a termination criterion of $\hat{\sigma}_{(p)}$ equal to or less than .32, the validity coefficient for a sample size of 50 was .928. In attaining the termination criterion, the minimum, average, and maximum number of items used in evaluation were 8, 12.2, and 16, respectively. Analogous interpretations apply to the remaining rows of data in the table. In the simulation of the tailoring of the VSAT, *Item Bank C*, it was found that the validity coefficient for the 80-item raw total score was .949, which was the same as that reported in Table 3 for tailored tests of an average length of 27.6 items. On the average, 65% of the original test may be considered unnecessary for examinees to take, since comparable validity in Bayesian tailored tests can be obtained with this substantial reduction in test length or items used.

## Discussion

In the parlance of the factor analyst, the validity coefficient used here is similar to a factor structure coefficient or the correlation between the variable and the factor of ability. In the present context, the novel usage

Table 3

Validity Coefficients and Pertinent Data for Bayesian Tailored Testing with Three Item Banks

| Item | Termination | Number of Items | | | Validity | Sample |
|---|---|---|---|---|---|---|
| Bank | Criterion | Minimum | Average | Maximum | Coefficient | Size |
| A | $\hat{\sigma}_{(P)} \leq .32$ | 8 | 12.2 | 16 | .928 | 50 |
| | $\hat{\sigma}_{(p)} \leq .25$ | 14 | 18.4 | 22 | .956 | 50 |
| B | $\hat{\sigma}_{(p)} \leq .32$ | 8 | 11.5 | 14 | .927 | 50 |
| | $\hat{\sigma}_{(p)} \leq .25$ | 12 | 17.4 | 21 | .948 | 50 |
| C | $\hat{\sigma}_{(p)} \leq .25$<br>or p = 30 | 14 | 27.6 | 30 | .949 | 100 |

of the word "variable" should not obscure the direct analogy. However, there is an important methodological distinction to be made. The logistic model allows one to avoid difficulty factors. For example, if we were to intercorrelate the items from any of the item banks--A, B, or C--the correlation matrix would form a simplex. In a simplex, the magnitudes of the correlations between pairs of items $\begin{pmatrix}\text{increase}\\\text{decrease}\end{pmatrix}$ as the disparities in the item difficulties $\begin{pmatrix}\text{decrease}\\\text{increase}\end{pmatrix}$. In factor analyzing a matrix of this nature, a plurality of factors is possible even though the basic source of the data is explicitly unidimensional. The model, then, circumvents a problem in factor analysis that has been viewed by several investigators (Ferguson, 1941; Gibson, 1960; Green, 1952) with some concern.

If the logistic model fits empirical data, the inferences made from simulation studies are applicable to those empirical situations where the constituent items of an item bank correspond in terms of their parameters to those that have been simulated. As seen above, one may fix ability by choosing a person at random or by selecting a gaussian random number. As far as the model is concerned, the operations are identical since underlying ability is assumed to be N(0,1). Thereafter, the model makes explicit the probability distributions from which one samples the response vectors for items of specified parameters. The upshot is that, if the model obtains for empirical data, simulated random samples do not differ in any critical way from empirical random samples.

Obviously, one can assess model capabilities independent of the determination of the fit of the model to empirical data. The question of model capabilities is, therefore, the more basic since, if the model does not show sufficient promise, tests of empirical fit are superfluous. Given sufficient promise, the question of empirical fit, then, becomes important. Methods of assessing the fit of the model to empirical data are discussed by Birnbaum (1968). While the results of this investigation would underscore the importance of the methods, no attempt will be made here to examine them.

Now Item Banks A and B were presented as idealized examples; but, if one has the specific objective of tailoring tests, reasonably similar banks might be achieved in practice. On the other hand, Item Bank C has (admitting to minimal errors in estimation due to the large sample size) its counterpart in the form of an extant conventional test, the VSAT. The findings with regard to Item Bank C indicated that if the VSAT were used in Bayesian tailored testing applications, 65% of the test would be unnecessary for the average examinee to take. However, if we look at the results for Item Banks A and B, an average of 27.6 items in relation to 18.4 and 17.4 items for evaluation with comparable validity shows that there is room for improvement. The improvement would be realized by designing item banks for the specific purpose of tailored testing.

The VSAT was designed for the specific purpose of univariate selection where the selection ratio is low. In other words, the test was constructed to minimize the errors of measurement in the range of high scores. To accomplish this purpose, items of higher than average difficulty were more frequently selected to comprise the test. In a tailored testing context, however, it is advisable to have a distribution of item difficulties which extends uniformly through the full range of difficulty (Urry, 1970). The specific purpose of tailored testing, then, is best served by a different item selection technique than that used in the construction of the VSAT.

Notice, further, that we have not utilized the Bayesian estimation procedures to the fullest extent. In most testing applications we begin under the tacit assumption that we know nothing about the examinees. In the majority of cases, this state of ignorance need not be assumed. Further assessments of model capabilities in regard to Bayesian estimation procedures should explore this attractive feature. The possibility exists that evaluation sequences could be further reduced in terms of average number of items while a satisfactory level of validity is maintained.

# References

Birnbaum, A. Part 5. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, Statistical theories of mental test scores, Reading, Mass.: Addison-Wesley, 1968.

Brogden, H. E. Latent ability and the structure of tests. Mimeographed. West Lafayette, Ind.: Purdue University, 1971.

Ferguson, G. A. The factorial interpretation of test difficulty. Psychometrika, 1941, 6, 323-329.

Gibson, W. A. Nonlinear factors in two dimensions. Psychometrika, 1960, 25, 381-392.

Green, B. F. Latent structure analysis and its relation to factor analysis. Journal of the American Statistical Association, 1952, 22, 71-76.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020.

Lord, F. M. A theoretical study of the measurement effectiveness of flexilevel tests. Research Bulletin 71-6. Princeton, N. J.: Educational Testing Service, 1971.

Owen, R. J. A Bayesian approach to tailored testing. Research Bulletin 69-92. Princeton, N. J.: Educational Testing Service, 1969.

Urry, V. W. A Monte Carlo investigation of logistic mental test models. Unpublished doctoral dissertation. Purdue University, 1970.

# Appendix

The order and preparation of the input specific to the computer programs are described here. Also, source listings are provided for both programs. Case I, where chance success is not effective, corresponds to program BAZES2. The particular designation is used to indicate that the items used in the Bayesian estimation of ability have but two parameters since the $c_i$ are null. Similarly, program BAZES3 corresponds to Case II where chance success on the items is effective or the $c_i$ are non-null.

The author wishes to acknowledge the able assistance of Jerry W. Edwards of the Bureau of Testing. His programming of FUNCTION ERF, which is used in both programs, was of considerable aid.

Order and Preparation of Cards Specific to the Programs

(1) Title Card (Format #200)

TITLE Col. 1-70 Any alphanumeric title for the identification of individual evaluation sequences.

NVF Col. 73-76 Number of variable format cards. Maximum: 5

NCASES Col. 77-80 Number of individuals to be evaluated.

(2) Problem Card (Format #74)

NAME Col. 1-40 Name of the item bank.

ITERM Col. 64 Specification of termination of criteria. Indicate:

1 If the evaluation sequences are to be terminated after a given number of item presentations.

2 If the evaluation sequences are to be terminated after an allowable value of the standard error of estimate has been attained.

3 If either of the conditions stated for 1 or 2 obtain.

IUSE Col. 67-68 Maximum number of items to be presented. Must be specified if Col. 64 contains a 1 or 3. Cannot exceed the number of items in the bank.

EPSILON Col. 69-76 Maximum allowable value for the standard error of estimate. Use F8.4. Must be specified if Col. 64 contains a 2 or 3. Suggested range: .4 to .2.

NITEMS Col. 77-80 Number of items in the item bank. Maximum: 200.

(3) Item Parameter Cards (Format #31)

(a) Item discriminatory powers ($a_i$)

Ten per card:

| | | |
|---|---|---|
| Col. 1-8 | $a_1$ | F8.4 |
| Col. 9-16 | $a_2$ | F8.4 |
| . . . . . . | | |
| Col. 73-80 | $a_{10}$ | F8.4 |

Repeat on the required number of cards. Maximum: 200 parameters or 20 cards.

(b) Item difficulties ($b_i$)

Ten per card:

| | | |
|---|---|---|
| Col. 1-8 | $b_1$ | F8.4 |
| Col. 9-16 | $b_2$ | F8.4 |
| . . . . . . | | |
| Col. 73-80 | $b_{10}$ | F8.4 |

Repeat on the required number of cards. Maximum: 200 parameters or 20 cards.

Necessary only for program BAZES3

(c) Probability of chance success on the items ($c_i$)

Ten per card:

| | | |
|---|---|---|
| Col. 1-8 | $c_1$ | F8.4 |
| Col. 9-16 | $c_2$ | F8.4 |
| . . . . . . | | |
| Col. 73-80 | $c_{10}$ | F8.4 |

Repeat on the required number of cards. Maximum: 200 parameters or 20 cards.

(4) Variable Format Card(s)

A maximum of five cards can be used to describe the data. Use columns 1-80. Punch a regular FORTRAN format statement omitting the word "FORMAT." Begin with "(I4," to accommodate an individual identification number, and use "I1" to input each binary response.

(5) Input Data (Variable Format)

The input data consist of an identification number and binary responses for each case as described on the variable format card(s).

```
      PROGRAM BAZES2(INPUT, OUTPUT, TAPE5=INPUT, TAPE6=OUTPUT)
      DIMENSION A(200),B(200),D(200),ALPHA(200),IR(200),IELIM(200) ,
     X          ISUB(200),R(200),ER(200),FMT(100),TITLE(7),NAME(4)
      PI=3.141592653589793
      C1 = 1.0/SQRT(PI)
      C2 = 2.0/SQRT(PI)
      C3 = SQRT(2.0/PI)
      SIGN = -1.0
200   READ(5,201) (TITLE(I),I=1,7),NVF,NCASES
201   FORMAT (7A10,2X,2I4)
      IF(EOF,5) 300,301
301   WRITE(6,70)
70    FORMAT(1H1,///////////////)
      WRITE(6,71)
71    FORMAT(1H0,47X,35HBAYESIAN ABILITY ESTIMATION PROGRAM/54X,
     X23HTWO ITEM PARAMETER CASE//////)
      WRITE(6,72)
72    FORMAT(1H0,56X,17HBUREAU OF TESTING /////)
      WRITE(6,73)
73    FORMAT(1H0,52X,24HUNIVERSITY OF WASHINGTON ////////)
      READ(5,74)NAME,ITERM,IUSE,EPSILON,NITEMS
74    FORMAT(4A10,22X,I2,I4,F8.4,I4)
      WRITE(6,75)NAME,NITEMS
75    FORMAT(1H0,25X,10HITEM BANK ,5X,4A10////26X,12HTOTAL ITEMS ,4X,
     XI4////)
      IF(ITERM-2) 76,77,78
76    EPSILON=.0
      GO TO 79
77    IUSE=NITEMS
78    CONTINUE
79    WRITE(6,80) IUSE,EPSILON
80    FORMAT(1H0,25X,20HTERMINATION CRITERIA//30X,16HNUMBER OF ITEMS ,
     XI10//30X,15HSTANDARD ERROR ,3X,F10.4)
      READ (5,31) (A(I),I=1,NITEMS)
      READ (5,31) (B(I),I=1,NITEMS)
31    FORMAT(10F8.4)
      WRITE(6,84)
84    FORMAT(1H1)
      CALL VARFMT (NVF,FMT)
      NK = 0
100   NK = NK + 1
      IPRES = 0
      IF (NK-NCASES) 101,101,200
101   WRITE (6,28) (TITLE(I),I=1,7)
28    FORMAT(1H1,30X,7A10////)
      READ(5,FMT) ID,(IR(I),I=1,NITEMS)
      ABLE=0.0
      VAR=1.0
      SEE=SQRT(VAR)
      WRITE(6,81) ID,ABLE,VAR
81    FORMAT(1H0,10X,15HEXAMINEE NUMBER,I10//16X,15HPRIOR ESTIMATES//21X
     X,8HABILITY ,F10.5/21X,9HVARIANCE ,F9.5//)
      WRITE(6,82)
82    FORMAT(1H0,15X,4HITEM,24X,9HITEM BANK,24X,7HABILITY,22X,
     X14HSTANDARD ERROR/12X,12HPRESENTATION,21X,6HNUMBER,26X,8HESTIMATE,
     X22X,11HOF ESTIMATE//)
```

(1)

```
        DO 20 I=1,NITEMS
 20     IELIM(I) = 0
 50     NLEFT = 0
        IF(IPRES.GE.IUSE) GO TO 100
        IF(EPSILON.GE.SEE) GO TO 100
        DO 13 I=1,NITEMS
        IF (I-IELIM(I)) 12,13,12
 12     NLEFT = NLEFT + 1
        ISUB(NLEFT) = I
 13     CONTINUE
        IF(NLEFT) 100,100,14
 14     DO 10 I=1,NLEFT
        ITEM = ISUB(I)
        R(I) = 1.0/(A(ITEM)*A(ITEM)) + VAR
        D(I) = (B(ITEM)-ABLE)/SQRT(2.0*R(I))
        ER(I)=ERF(D(I))
 10     ALPHA(I) = R(I)*EXP(2.0*D(I)*D(I))*(1.0-ER(I)*ER(I))
        RMIN = ALPHA(1)
        JSUB = 1
        DO 1 J=1,NLEFT
        IF(RMIN - ALPHA(J)) 1,1,2
 2      RMIN = ALPHA(J)
        JSUB = J
 1      CONTINUE
        ITEM = ISUB(JSUB)
        RITEM = R(JSUB)
        DITEM = D(JSUB)
        IELIM (ITEM) = ITEM
        IPRES = IPRES + 1
        TEM = DITEM*DITEM
        F1 = C2*(1.0/(1.0+(1.0/(A(ITEM)*A(ITEM))*(1.0/VAR))))
        IF(IR(ITEM)) 3,3,4
 3      PER = 1.0+ER(JSUB)
        ABLE = ABLE-C3*(VAR/SQRT(RITEM))*EXP(SIGN*TEM)/PER
        F2 = 1.0/((EXP(TEM)*PER)**2)
        F3 = C1+DITEM*EXP(TEM)*PER
        GO TO 5
 4      SER = 1.0-ER(JSUB)
        ABLE = ABLE+C3*(VAR/SQRT(RITEM))*EXP(SIGN*TEM)/SER
        F2 = 1.0/((EXP(TEM)*SER)**2)
        F3 = C1-DITEM*EXP(TEM)*SER
 5      VAR = VAR *(1.0-F1*F2*F3)
        SEE=SQRT(VAR)
        WRITE(6,83)IPRES,ITEM,ABLE,SEE
 83     FORMAT(1H0,15X,I4,26X,I4,26X,F8.4,25X,F8.4)
        GO TO 50
 300    STOP
        END
        FUNCTION ERF(X)
 *      ERROR FUNCTION FOR) 0 I X - 3.95, MACCLAURIN SERIES
 *      FOR X * 10, ERF(X)=1.0
 *      ACCURACY PARAMETER IS .1E-10
        ERF=0.0$IF(X.EQ.0.0)RETURN$ES=SIGN(1.0,X)$XX=E=Y=ABS(X)$J=1$IF(Y.L
       XT.3.96)GOTO1$ERF=1.0$RETURN
      1 S=F=1$DO2I=1,60$S=S*(-1)$F=F*I$J=J+2$XX=XX*Y*Y
        T=XX/(J*F)$IF(T.LT..1E-10)GOTO3$E=E+T*S
```

(2)

```
    2 CONTINUE
    3 ERF=ABS(E)*2.0/1.77245385090552*ES$RETURN$END
      SUBROUTINE VARFMT(NVF,FMT)
C     NVF= NUMBER OF VARIABLE FORMAT CARDS
      DIMENSION FMT(1)
      NVF=NVF*20
      READ(5,470) (FMT(I),I=1,NVF)
  470 FORMAT(20A4)
      WRITE(6,471)(FMT(I),I=1,NVF)
 471  FORMAT(1H0,7H FORMAT/(10X,20A4))
      RETURN
      END
```

(3)

```
      PROGRAM BAZES3(INPUT, OUTPUT, TAPE5=INPUT, TAPE6=OUTPUT)
      DIMENSION A(200),B(200),C(200),D(200),BETA(200),IR(200),
     X          ISUB(200),R(200),ER(200),FMT(100),TITLE(7),NAME(4),
     X          IELIM(200)
      PI=3.141592653589793
      C1 = 1.0/SQRT(PI)
      C2 = 2.0/SQRT(PI)
      C3 = SQRT(2.0/PI)
      C4 = 1.0/SQRT(2.0*PI)
      SIGN = -1.0
  200 READ(5,201) (TITLE(I),I=1,7),NVF,NCASES
  201 FORMAT (7A10,2X,2I4)
      IF(EOF,5) 300,301
  301 WRITE(6,70)
  70  FORMAT(1H1,///////////////)
      WRITE(6,71)
  71  FORMAT(1H0,47X,35HBAYESIAN ABILITY ESTIMATION PROGRAM/53X,
     X25HTHREE ITEM PARAMETER CASE//////)
      WRITE(6,72)
  72  FORMAT(1H0,56X,17HBUREAU OF TESTING /////)
      WRITE(6,73)
  73  FORMAT(1H0,52X,24HUNIVERSITY OF WASHINGTON ////////)
      READ(5,74)NAME,ITERM,IUSE,EPSILON,NITEMS
  74  FORMAT(4A10,22X,I2,I4,F8.4,I4)
      WRITE(6,75)NAME,NITEMS
  75  FORMAT(1H0,25X,10HITEM BANK ,5X,4A10////26X,12HTOTAL ITEMS ,4X,
     XI4////)
      IF(ITERM-2) 76,77,78
  76  EPSILON=.0
      GO TO 79
  77  IUSE=NITEMS
  78  CONTINUE
  79  WRITE(6,80) IUSE,EPSILON
  80  FORMAT(1H0,25X,20HTERMINATION CRITERIA//30X,16HNUMBER OF ITEMS ,
     XI10//30X,15HSTANDARD ERROR ,3X,F10.4)
      READ (5,31) (A(I),I=1,NITEMS)
      READ (5,31) (B(I),I=1,NITEMS)
      READ (5,31) (C(I),I=1,NITEMS)
  31  FORMAT(10F8.4)
      WRITE(6,84)
  84  FORMAT(1H1)
      CALL VARFMT (NVF,FMT)
      NK = 0
  100 NK = NK + 1
      IPRES = 0
      IF (NK-NCASES) 101,101,200
  101 WRITE (6,28) (TITLE(I),I=1,7)
  28  FORMAT(1H1,30X,7A10////)
      READ(5,FMT) ID,(IR(I),I=1,NITEMS)
      ABLE=0.0
      VAR=1.0
      SEE=SQRT(VAR)
      WRITE(6,81) ID,ABLE,VAR
  81  FORMAT(1H0,10X,15HEXAMINEE NUMBER,I10//16X,15HPRIOR ESTIMATES//21X
     X,8HABILITY ,F10.5/21X,9HVARIANCE ,F9.5//)
      WRITE(6,82)
```

(4)

```
     FORMAT(1H0,15X,4HITEM,24X,9HITEM BANK,24X,7HABILITY,22X,
    X14HSTANDARD ERROR/12X,12HPRESENTATION,21X,6HNUMBER,26X,8HESTIMATE,
    X22X,11HOF ESTIMATE//)
     DO 20 I=1,NITEMS
     IELIM(I) = 0
     NLEFT = 0
     IF(IPRES.GE.IUSE) GO TO 100
     IF(EPSILON.GE.SEE) GO TO 100
     DO 13 I=1,NITEMS
     IF (I-IELIM(I)) 12,13,12
     NLEFT = NLEFT + 1
     ISUB(NLEFT) = I
     CONTINUE
     IF(NLEFT) 100,100,14
     DO 10 I=1,NLEFT
     ITEM = ISUB(I)
     R(I) = 1.0/(A(ITEM)*A(ITEM)) + VAR
     D(I) = (B(ITEM)-ABLE)/SQRT(2.0*R(I))
     ER(I)=ERF(D(I))
     SER = 1.0-ER(I)
     CITEM = C(ITEM)
     C5 = 1.0-CITEM
     F1 = 1.0-(SER/2.0)
     F2 =(CITEM+(C5/2.0)*SER)*EXP(2.0*D(I)*D(I))
     BETA(I) = (1.0/C5)*(R(I)/VAR)*F1*F2*(1.0/(1.0+CITEM*F1))
     RMIN = BETA(1)
     JSUB = 1
     DO 1 J=1,NLEFT
     IF(RMIN - BETA(J)) 1,1,2
     RMIN = BETA(J)
     JSUB = J
     CONTINUE
     ITEM = ISUB(JSUB)
     RITEM = R(JSUB)
     DITEM = D(JSUB)
     IELIM (ITEM) = ITEM
     IPRES = IPRES + 1
     TEM = DITEM*DITEM
     IF(IR(ITEM)) 3,3,4
     PER = 1.0+ER(JSUB)
     ABLE = ABLE-C3*(VAR/SQRT(RITEM))*EXP(SIGN*TEM)/PER
     F1 = C2*(1.0/(1.0+(1.0/(A(ITEM)*A(ITEM))*(1.0/VAR))))
     F2 = 1.0/((EXP(TEM)*PER)**2)
     F3 = C1+DITEM*EXP(TEM)*PER
     GO TO 5
     SER = 1.0-ER(JSUB)
     CITEM = C(ITEM)
     C5 = 1.0-CITEM
     F4 = (1.0/(CITEM+(C5/2.0)*SER))
     ABLE=ABLE+C4*C5*F4*EXP(SIGN*TEM)*(VAR/SQRT(RITEM))
     F1 = C5*C1*F4
     F2 = (VAR/RITEM)*EXP(2.0*SIGN*TEM)/SER
     F3 = ((C1-DITEM*EXP(TEM)*SER)-(CITEM*C1*F4))
5    VAR = VAR *(1.0-F1*F2*F3)
     SEE=SQRT(VAR)
     WRITE(6,83)IPRES,ITEM,ABLE,SEE
```

```
 83     FORMAT(1H0.15X,14,26X,I4,26X,F8.4,25X,F8.4)
        GO TO 50
 300    STOP
        END
        FUNCTION ERF(X)
 *      ERROR FUNCTION FOR) 0 I X - 3.95, MACCLAURIN SERIES
 *      FOR X * 10, ERF(X)=1.0
 *      ACCURACY PARAMETER IS .1E-10
        ERF=0.0$IF(X.EQ.0.0)RETURN$ES=SIGN(1.0,X)$XX=E=Y=ABS(X)$J=1$IF(Y.L
       XT.3.96)GOTO1$ERF=1.0$RETURN
      1 S=F=1$DO2I=1,60$S=S*(-1)$F=F*I$J=J+2$XX=XX*Y*Y
        T=XX/(J*F)$IF(T.LT..1E-10)GOTO3$E=E+T*S
      2 CONTINUE
      3 ERF=ABS(E)*2.0/1.77245385090552*ES$RETURN$END
        SUBROUTINE VARFMT(NVF,FMT)
 C      NVF= NUMBER OF VARIABLE FORMAT CARDS
        DIMENSION FMT(1)
        NVF=NVF*20
        READ(5,470) (FMT(I),I=1,NVF)
    470 FORMAT(20A4)
        WRITE(6,471)(FMT(I),I=1,NVF)
   471  FORMAT(1H0,7H FORMAT/(10X,20A4))
        RETURN
        END
```