SOME LIKELIHOOD FUNCTIONS FOUND IN TAILORED TESTING

FREDERIC M. LORD
Educational Testing Service

This brief note discusses some peculiar likelihood functions encountered while administering the Broad-Range Tailored Test of Verbal Ability to simulated examinees. Other workers have doubtless encountered similar problems.

Samejima (1973) shows that when the item parameters are known, there may be no finite ability level $\hat{\theta}$ that maximizes the likelihood function. Also, that the likelihood function may have more than one (local) maximum.

Barnett (1966) states "Given a single sample of observations ... [r]egularity conditions ... are no guarantee that a *single* root of the likelihood equation will exist for this sample. In fact, there will often exist multiple roots, corresponding to multiple relative maxima of the likelihood function, even if the regularity conditions are satisfied."

Huzurbazar (see Kendall & Stuart, 1973, sections 18.11-18.12) showed under regularity conditions that ultimately, as the number of observations becomes large, there is a *unique* consistent maximum likelihood estimator. His regularity conditions would apply if the test were composed of items with identical ICC. His conditions would be violated otherwise, but it should be possible to extend his proof to cover a reasonable set of regularity conditions for the present problem.

To have a large number of observations, we would need to administer a large number of test items. When the number of items is not large, and especially if the test is too hard for some individuals, we may expect $\hat{\theta}_a = -\infty$ occasionally. An examinee who makes unlucky guesses and scores below the chance level is, not unreasonably, likely to get an estimated ability of $\hat{\theta} = -\infty$. Such an estimate would presumably be corrected if a sufficiently large number of additional test items were administered to him.

In the study on a Broad-Range Tailored Test of Verbal Ability, many tens of thousands of simulated examinees took various simulated tailored tests. Items with known ICC were administered one at a time to each individual examinee. After each item was adminsitered, an approximation to the maximum likelihood estimate $\hat{\theta}$ of his ability was computed, based on all his responses up to that point.

When the examinee has wrong answers but no right answers, $\hat{\theta} = -\infty$. When he has right answers but no wrong answers, $\hat{\theta} = +\infty$. When he has both right and wrong answers, there is usually no difficulty in finding a finite $\hat{\theta}$. An occasional difficulty resolves itself as more items are administered. It is very rare to have any problem after the first ten or fifteen items, since by then the item difficulty is usually tolerably well tailored to the examinee's ability.

The present study investigates the case of simulated examinee T94 for whom there were unusual difficulties in obtaining a finite $\hat{\theta}$. Table 1 describes the first 23 items administered to him, shows his response to each item (1 = right, 0 = wrong), and gives $\hat{\theta}$, the maximum likelihood estimate of his ability based on his responses to items already administered.

Examinee T94 is really a very low ability examinee—his true θ is actually -2.9. Furthermore, the first items administered to him were very difficult items ($b_i > 1.35$) which he would have no chance at all of answering correctly except by guessing. By lucky guessing, he nevertheless got 6 items right out of the first 12.

If c_i were .20 for each of these items, the chance of a score as good or better than 6 solely by guessing is less than .02. The maximum likelihood estimates of the examinee's ability based on his performance on these first twelve items range from 1.6 to 2.2, as shown in the last column of the table.

His guessing on the next seven items was uniformly unsuccessful. All items through item 17 were difficult, with $b_i > 1.35$. His performance on these 17 difficult items earned him an ability estimate of $\hat{\theta} = 1.2$.

Item 18 was an easier item, $b_{18} = .65$. I suggest that the following rationalizations provide a correct explanation of the θ subsequently obtained.

The examinee has answered correctly 6 items with $b_i > 1.35$ and has failed 12 items including one with $b_i = .65$. The last failure suggests that θ is low and that earlier correct responses were due to lucky guessing. If θ is low, all items so far administered are too difficult for the examinee and are of no use, even for placing a lower bound on his ability level. When an examinee has given only wrong responses and lucky random guesses, his estimated ability should be $\hat{\theta} = -\infty$.

When the examinee answers item 20 ($b_{20} = -.83$) correctly, it is now plausible to assume that his ability lies between $-.83$ and .65 (.65 being the difficulty level of item 18, which he answered incorrectly). The maximum likelihood estimate turns out to be $\hat{\theta} = -.4$.

Research reported in this paper has been supported by grant GB-41999 from National Science Foundation.

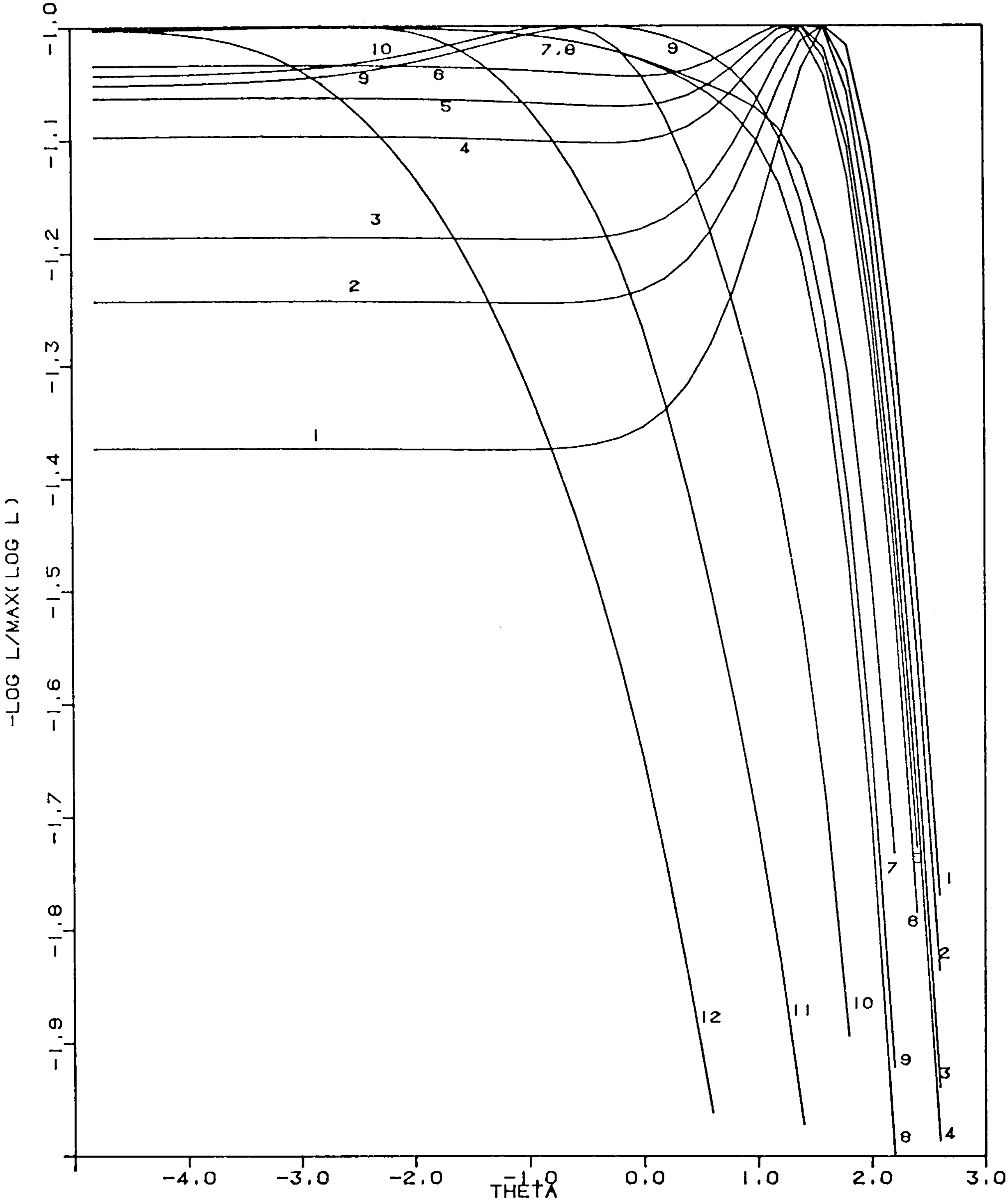

Figure 1. Standardized likelihood function for examinee no. T94, $\theta = -2.9$.

TABLE 1

Successive Estimates of Ability for Examinee T94

Item no.	Curve no. in Fig. 1	Item Parameters a	b	c	Examinee's response	Number of right answers	Log likelihood* at $\hat{\theta}$	Estimated ability** $\hat{\theta}$
1		.61	2.20	.19	0	0		–
2		2.05	1.74	.18	1	1		2.2
3		1.48	2.51	.17	0	1		1.9
4		1.89	1.96	.20	0	1		1.6
5		1.93	1.89	.24	1	2		1.8
6		2.21	1.73	.21	1	3		2.0
7		1.57	1.76	.07	0	3		1.8
8		1.68	1.40	.15	1	4		1.8
9		1.42	1.36	.13	0	4		1.7
10		1.27	1.65	.28	1	5		1.7
11		1.56	1.49	.19	0	5		1.6
12	1	1.34	1.54	.19	1	6	–7.7	1.6
13	2	1.07	1.52	.20	0	6	–8.7	1.6
14	3	1.31	1.89	.09	0	6	–9.2	1.4
15	4	.93	1.35	.20	0	6	–10.1	1.4
16	5	1.02	1.98	.21	0	6	–10.7	1.4
17	6	1.03	1.88	.13	0	6	–11.1	1.2
18	7	1.24	.65	.20	0	6	–11.7	– ∞
19	8	2.00	1.27	.10	0	6	–11.8	– ∞
20	9	.88	–.83	.33	1	7	–12.3	–.4
21	10	2.10	.05	.21	0	7	–12.6	–.8
22	11	1.37	–1.49	.15	0	7	–13.3	-2.6
23	12	1.10	–2.84	.24	0	7	–13.6	– ∞

*Not computed for $n < 12$.

**For $n = 2,3,\ldots,11$, the listed $\hat{\theta}$ is an approximate value determined numerically. For $n > 11$, the listed was $\hat{\theta}$ was read from values of the log likelihood tabulated at intervals of .2 along the θ scale.

Subsequent failures on items 21 and 22 lower this estimate to -.8 and then to -2.6. When the examinee finally fails an item with $b_i = -2.84$, it now appears that all earlier correct answers were due to lucky guessing and that all items so far administered were too difficult for this examinee. The situation is much the same as the situation after the answer to item 18, already discussed. Again, not unreasonably, $\hat{\theta} = -\infty$.

In this testing, only the very last item was of appropriate difficulty for the examinee, whose true ability was $\theta = -2.9$. All but the last two items were very much too hard. He answered both the last two items incorrectly. Thus, it is only to be expected that his final ability estimate is $\hat{\theta} = -\infty$. Administration of further items of appropriate difficulty would quickly correct this estimate.

The likelihood functions used to obtain most of the successive $\hat{\theta}$ discussed above are shown in Figure 1. The code numbers identifying the curves are given in Table 1. In order to get them all on the same graph, each likelihood function is divided by its maximum value, so that the maxima of the normalized curves all fall on the top boundary of the figure. These curves, together with the discussion given above, seem to explain the anomalous values of $\hat{\theta}$. When enough responses have been obtained to indicate a lower limit to the examinee's ability, then finite ability estimates will be obtained.

REFERENCES

Barnett, V. D. Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. *Biometrika*, 1966, *53*, 151-165.

Kendall, M. G. and Stuart, A. *The advanced theory of statistics.* New York: Hafner, Vol. 1, 1969; Vol. 2, 1973.

Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. *Psychometrika*, 1973, *38*, 221-233.