THE WEIBULL DISTRIBUTION AND A PRACTICAL ANALYSIS OF MENTAL TEST SCORES

SHIGEO KASHIWAGI

Japan National Railroads

We propose to apply the Weibull distribution to the analysis of mental test scores. As the first place, we review the distribution theory of test scores by Lord (1952) and suggest the relationship between the theory and the shape parameter of this distribution.

First of all, we describe the simple Weibull distribution and the discussion concerning the estimation of the shape parameter is made. Especially, the emphasis is placed on the method by means of the Weibull Probability Paper (wpp), and it is pointed out that the method is very useful and recommendable in view of being accurate practically.

Secondly, we discuss the concept of measurability of test, and make comment on the difference between the discriminability of test and the measurability. And further, the mixed or composite Weibull distribution is suggested to be useful under a certain condition.

Finally, several numerical examples are presented and we discuss some of the important aspects of our method based on the Weibull distribution.

INTRODUCTION

The Weibull distribution was proposed by a Swedish statistician W. Weibull (1951). Following him, many statisticians and researchers studied the distribution and recognized it of great use. Today, the distribution is widely used in natural science. This paper is concerned with the application of the distribution to the analyses of mental test scores.

Lord (1952) proposed a model of the distribution of test scores. He demonstrated the frequency distribution of true scores to be determined by the latent trait and the test characteristic curve (tcc), and showed the shape of the tcc by the discriminating power (or discriminability) and the difficulty of test. In other words, under a given distribution of the latent trait, the tcc will play an important role in determining the shape of the distribution of true scores. If the converse is true, we may get useful informations on the discriminating power and difficulty of test only through the estimation of the shape of the distribution of true scores.

The mathematical concept of the failure rate has long been discussed in the field of life testing and reliability theory. The concept discriminates distributions effectively. We shall call it the conditional ratio of difficulty (crd).

Lord assumed various kinds of shapes of the distributions of true scores. The Weibull distribution may be very appropriate in that its crd meets his requirement very well. Some of distributions, e.g., log-normal, gamma, and, negative hypergeometric distribution (Keats & Lord, 1962), look like the Weibull distribution. However, unlike the Weibull distribution, they lack the simple and remarkable relationship between the crd and the shape parameter of the distribution.

From the practical point of view, at the
The Weibull Distribution and a Practical Analysis of Mental Test Scores

present time, the negative hypergeometric distribution is said to be most satisfactory in the sense of good of fitness. However, when a distribution of test scores is extremely skewed, the negative hypergeometric distribution often fail to close fit. The trouble may be improved by using the Weibull distribution.

THE CONDITIONAL RATIO OF DIFFICULTY AND THE DISTRIBUTION OF TEST SCORES

When \( f(x) \) denotes the density function of a distribution of test scores, the cumulative distribution \( F(x) \) is written

\[
F(x) = \int_0^x f(t) dt, \quad (1)
\]

where

\[
\int_0^\infty f(t) dt = 1
\]

is assumed. Ordinarily, in the test theory, we use

\[
F(x) = \int_{-\infty}^x f(t) dt
\]

rather than (1). However, we here transform the test scores so as to satisfy (1).

Let us introduce the following function:

\[
\lambda(x) = \frac{f(x)}{1-F(x)} = \frac{1-F(x)}{1-F(x)} \quad (2)
\]

As we pointed out, (2) is called the failure rate in the field of reliability theory. Lord (1968; p. 409) treats \( -f(x) \) as a measure related to the discriminating power and the difficulty of a test under a given level of ability. Here, we call (2) the crd. From (2) we get

\[
1-F(x) = \exp \left( -\int_0^x \lambda(t) dt \right). \quad (3)
\]

When the crd is constant, the distribution yields the exponential one. This is the simplest case. However, some distributions have rather complicated crds. For example, the crd of the negative hypergeometric distribution will be the case. On the contrary, the crd of the Weibull distribution is simple and the value of the shape parameter determines the distribution of the crd definitely. The explanation concerning the Weibull distribution will be given in the following section.

THE WEIBULL DISTRIBUTION

The Weibull distribution has the shape, scale, and location parameters. However, the estimation of the location parameter will not be discussed hereafter, because the estimation of it is not fundamental problem in test theory.

\[
\lambda(x) = \frac{m}{a} x^{m-1}, \quad x \geq 0; \quad 0, \quad x \leq 0. \quad (4)
\]

Here, \( m \) and \( a \) represent the shape and scale parameters, respectively. This distribution of the crd is determined by the value of \( m \). When \( m \) is smaller than unity the distribution is smoothed L shaped, and when \( m \) is larger than unity the distribution is J shaped. The distribution of the crd is horizontal when \( m \) is equal to unity. This is the simplest case and we call it the exponential distribution. And, from (3) and (4) we get

\[
1-F(x) = e^{-\frac{x^m}{a}}, \quad x \geq 0; \quad 1, \quad x \leq 0. \quad (5)
\]

and

\[
f(x) = \frac{m}{a} x^{m-1} e^{-\frac{x^m}{a}}, \quad x \geq 0; \quad 0, \quad x \leq 0. \quad (6)
\]

The value of the shape parameter \( m \) is invariant with the value of the scale parameter. In other words, the estimation of \( m \) may be independent of the value of the scale parameter \( a \). When we replace \( x/a^{1/m} \) by \( X \), the equation (5) will be

\[
1-F(X) = e^{-X^m}, \quad X \geq 0; \quad 1, \quad X \leq 0. \quad (5')
\]

(5') is suggested the fact just mentioned above.

The \( k \)-th moment about the origin is
where

\[ \mu_1 = \int_0^\infty x^{m-1}e^{-\frac{x}{\alpha}}dx \]

is the gamma distribution. And the population mean and variance are respectively,

\[ \mu = \alpha \Gamma(m) \left( \frac{1}{m+1} \right) \]

\[ \sigma^2 = \alpha^2 \Gamma(m) \left( \frac{2}{m+1} - \Gamma\left( \frac{1}{m+1} \right) \right). \]

It should be noted that the estimation of the shape parameter of this distribution is tremendously simple and practical compared to the estimation of the shape of the gamma distribution having the similar shape of the distribution.

The Estimation of the Parameters

It is generally difficult to estimate three parameters of this distribution exactly. However, if the value of location parameter is assumed to be at zero, the estimation of the remaining two parameters may be easier and the estimated values will become more accurate. The maximum likelihood estimation is said to be most reliable and accurate. However, in this method, we will not attain a final solution without trial and error or successive iterative procedure. Although the objective estimation by the moment method is possible, the estimated values are not always useful because of their poor approximation. In the graphical procedure by means of the Weibull Probability Paper which we call hereafter wpp method, we must estimate the shape parameter by inspection. However, it is said that this is most practical and recommendable (Berrettioni, 1964).

We describe the principles of the maximum likelihood, the moment, and the wpp method in order. And we will discuss the analysis of the distribution of test scores by the wpp method.

1. The Principles of the Maximum Likelihood, the Moment, and the WPP Method

1-a. The maximum likelihood method

When the location parameter is zero, the maximum likelihood function \( L \) is written

\[ L(x_1, x_2, \ldots, x_n; m, \alpha) = \left( \frac{m}{\alpha} \right)^n \prod_{i=1}^n x_i^{m-1} \exp \left( -\frac{1}{\alpha} \sum_{i=1}^n x_i^m \right). \]

Taking logarithms of both sides of the equation,

\[ \ln L = n \ln m - n \ln \alpha + (m-1) \sum_{i=1}^n \ln x_i - \frac{1}{\alpha} \sum_{i=1}^n x_i^m \]

where \( \ln \) means the natural logarithm. Letting

\[ \frac{\partial (\ln L)}{\partial m} = \frac{\partial (\ln L)}{\partial \alpha} = 0, \]

we get

\[ \frac{1}{m} + \frac{1}{n} \sum_{i=1}^n \ln x_i - \frac{1}{\alpha} \sum_{i=1}^n x_i^m = 0, \]

where

\[ \sum_{i=1}^n x_i^m = n \alpha. \]

From (9), it may be possible to estimate \( m \) and \( \alpha \) in the trial and error method (cf. Cohen, 1965), or the iterative procedure based upon an initial approximation by the wpp method may be used (cf. Ohtsu, 1967).
1-b. The moment method

When the location parameter is zero, the expected value of the coefficient of variation \( E(\sigma / \mu_1) \) is a function of \( m \). From (7) and (8) we get

\[
E \left( \frac{\sigma}{\mu_1} \right) = \frac{\left( \Gamma \left( \frac{2}{m+1} \right) - \Gamma \left( \frac{1}{m+1} \right) \right)^{1/2}}{\Gamma \left( \frac{1}{m+1} \right)}.
\]

In practice, it is possible to estimate \( m \) by using the sample mean \( \bar{\mu}_1 \) and variance \( \bar{\sigma}^2 \) (Cohen, 1965).

1-c. The wpp method

From (5') we get

\[
\ln \ln \left( \frac{1}{1 - F(X)} \right) = m \ln X.
\]

If we let the left side of the equation be \( Y \),

\[ Y = m \ln X \quad (5'') \]

is obtained. According to (5'') the wpp is composed. In other words, \( \ln X \) is taken on the horizontal line and \( Y \) on the vertical one. As Fig. 1-a. shows, the upper horizontal line shows \( \ln X \) and the vertical one of the right side \( Y \). And further \( X \) is taken on the lower horizontal line and \( F(X) \) on the vertical one of the left side. If a given distribution is a Weibull one, the cumulative distribution must be on the line the slope of which is \( m \).

We will discuss the analysis of the distribution of test scores by the wpp method in the following section.

2. The Analysis of the Distribution of Test Scores by the WPP Method

Before describing in detail, we assume the following:

**Assumption:** A distribution of test scores is assumed to be one dimensional in the sense that the Weibull distribution fits with it irrespective of the direction measuring an assumed latent ability. Before the analysis starts, it should be determined which is to be analyzed, \( F(x) \) or \( 1 - F(x) \).

2-a. The linearization of \( F(x) \)

The following may be a practical procedure of the linearization of \( F(x) \):

Plot \( F(x) \) on a wpp. If it is \( J \) shaped, it should be moved to the right. And if it is like a normal orgive, it should be moved to the left. In other words, the trial and error method should be continued until the shape of \( F(x) \) will be close to a straight line.

2-b. The estimation of the parameters

The degree of the slope of the line on the wpp determines the value of the shape parameter. The slope of the line connecting two points with coordinates \((\ln X_1, Y_1)\) and \((\ln X_2, Y_2)\) is,

\[ m = \frac{Y_1 - Y_2}{\ln X_1 - \ln X_2}. \]

This is an estimated value of \( m \).

Next, the scale parameter \( \alpha \) may be estimated in the following way. If we let \( \ln \alpha \) be the intercept of \( \ln X \)-axis, the value

\[ \hat{\alpha} = e^{\ln X_1} \]

is an estimate of \( \alpha \). Or, from the intercept of \( Y \)-axis

\[ \hat{\alpha} = e^{Y_1} \]

is another estimate of \( \alpha \).

From (6) using \( \hat{m} \) and \( \hat{\alpha} \),

\[ f(\hat{x}) = \frac{\hat{m}}{\hat{\alpha}} \hat{x}^{\hat{m}-1} \exp \left( -\frac{\hat{x}^{\hat{m}}}{\hat{\alpha}} \right) \]

will be calculated. This may be an approximate estimation of the distribution of true scores. Here, \( \hat{x} \) shows an estimated test score on the wpp.

**Measurability and Difficulty**

Before discussion, we should like to point out that the following description could be applicable to the analyses of any kind of mental test scores. However, for easy explanation, the discussion will be confined
to the achievement type test in which test scores are counted by the number of right answers.

It is well known that the distributions of mental test scores give various kinds of shapes. They may be normally distributed or sometimes skewed extremely. Many psychometricians assumed various kinds of theoretical distributions for mental test scores in order to account for the actual distribution. Among them, Keats and Lord (1962) assumed that the negative hypergeometric distribution is most satisfactory in the sense of high degree of good of fitness. This was already pointed out.

In this paper, we propose to assume the Weibull distribution for the explanation of the distribution of mental test scores. Experience has indicated that the Weibull distribution fits well while other theoretical distributions do not fit so good (cf. Ex. 1 in the section of the Numerical Examples). Even if the simple Weibull distribution does not fit well, the mixed or composite Weibull one could be used. They may fit very well with almost all kinds of distributions for mental test scores.

We define the following new concept. It is measurability. If an efficient ordering of subjects is obtained by a given test, the test is said to be measurable. Needless to say, this may be a prerequisite concept for the theory and practice of psychological testing. As we mentioned, the crd may be defined as a powerful measure for the concept above. From the formula (4), it will be understood that the shape parameter \(m\) of the Weibull distribution plays an important role in determining the crd, and that the measurability of a test may be dependent upon the value of \(m\).

We define measurability independent of the direction of measuring mental trait. In other words, we do not make difference between the measurable test at the lower level, and at the high level, by the irrespectiveness of the direction to be measured. Here it should be noted the following. The measurable test for the higher scores (or the lower scores) means that it may make an efficient ordering of subjects especially for higher ability (or lower ability). By the definition of measurability, too large or small value of \(m\) may not be appropriate in the sense that the test which gives such a value is not measurable. Experience has indicated that we should take the value of \(m\) between 1.0 and 4.0. When the value of \(m\) for a given test is too large to both direction of higher and lower abilities, we will hardly be able to define the measurability for the test.

Suppose that a certain kind of tests are given to a sample, and suppose that they have various kinds of difficulties. When we have the tests each value \(m\) of which lies between 1.0 and 4.0 (i.e., we do not assume the tests to be extremely easy or difficult), the more difficult the test is, the smaller the value of \(m\) is. Here it should of course be noted that we assume all the tests to be measurable to the direction for the higher scores. We are thus able to estimate satisfactorily the difficulty of test only through the value of \(m\). This is our new measure for the difficulty of test. See Ex. 2, our application to Lord (1952)'s data, in the section of Numerical Examples.

**THE MIXED AND COMPOSITE WEIBULL DISTRIBUTIONS**

The following was already suggested by the present author. When the simple Weibull distribution does not fit well, the mixed or composite Weibull one should be used. They consist of two kinds of the Weibull distributions. And we differ hereafter the simple one from the mixed and from the composite ones.

The crd of the general mixed type distribution may be defined as follows. The distribution of mental test scores \(F(x)\) is assumed to be mixed by \(F_1(x)\) and \(F_2(x)\). If the ratios, \(F_1(x)/F(x)\) and \(F_2(x)/F(x)\), are shown by \(p_1\) and \(p_2\), \(F(x)\) is
The Weibull Distribution and a Practical Analysis of Mental Test Scores

\[ F(x) = p_1 F_1(x) + p_2 F_2(x) \]
\[ = \sum_{i=1}^{2} p_i F_i(x), \quad (11) \]
where
\[ p_1 + p_2 = \sum_{i=1}^{2} p_i = 1. \]

And the crd of the formula (11) is

\[ \lambda(x) = \frac{f(x)}{1 - F(x)} = \frac{\sum p_i f_i(x)}{1 - \left( \sum p_i F_i(x) \right)}. \quad (12) \]

The density function of the general composite type distribution is

\[ f(x) = \begin{cases} f_1(x); & \delta > x \geq r_1 \\ f_2(x); & x \geq \delta(> r_2) \end{cases} \quad (13) \]

where \( \delta \) is a parameter to divide the distribution into two parts in the following manner. That is

\[ F_1(\delta) = \int_{r_1}^{\delta} f_1(x)dx = F_2(\delta) = \int_{r_2}^{\delta} f_2(x)dx. \]

The crd is

\[ \lambda(x) = \begin{cases} f_1(x); & \delta > x \geq r_1 \\ f_2(x); & x \geq \delta(\geq r_2). \end{cases} \quad (14) \]

It is easy to fit an appropriate one of these type of the Weibull distributions with a given distribution of test scores by the method of wpp. And the fitting may also be made objectively, and then several different procedures were proposed (cf. e.g., Araki, 1968).

MEASURABILITY, DISCRIMINABILITY, AND, THE DESIGN OF THE DISTRIBUTION

As we have mentioned in the section of INTRODUCTION, Lord (1952) showed the frequency distribution of mental test scores to be determined by the discriminating power and difficulty of test when a distribution of latent trait is given. Keeping the difficulty parameter constant, the shape of the frequency distribution of mental test scores will be changed by the value of the discriminating power of the test. Let us take the tests composed of fifty percent difficulty, for instance. As the discriminating power of test increases, the shape of distribution of test scores will change from the distribution with high kurtosis to the rectangular distribution. And finally it will change into the U shaped one (cf. Lord, 1952). Lord demonstrated four illustrative examples of the distribution of relative true scores. The tests were composed of items of fifty percent difficulty, and their correlations with the latent trait were .20, .45, .65, and .80 respectively. We approximated the distributions by the Weibull ones. As the correlation increased the value of \( m \) decreased. And when the correlation was .80, the distribution became a mixed Weibull one. The value of \( m \) of a simple Weibull distribution was extremely small and the other one extremely large, and the sample was divided into two parts. Measurability was then hardly defined in each part of the sample. In other words, it was suggested that the increase of discriminability of test might not always mean the increase of measurability of test. These results suggest that, if we aim to obtain an efficient ordering of subjects, we should use a test of high measurability rather than a test of too much high discriminability. On the contrary, if we aim to divide the subjects into two parts, we should use a test of high discriminability.

The discussion above was confined to the test composed of fifty percent difficulty. However, we are not always able to assume such a kind of tests in the actual distributions of mental test scores, and we do not always have the simple and analytical relationships between the distribution of test scores and difficulty and discriminability of the test at the present stage. However, if a feasible simulation model, e.g., Birnbaum’s, for the relationships is used, and if we adopt the Weibull distribution for the distributions of test scores, the shape para-
The shape parameter $m$ of the Weibull distribution may be an integrated and interpretable measure for the shape of the distribution of test scores.

We proposed to take the value of the shape parameter $m$ of the Weibull distribution between 1.0 and 4.0, and this recommendation were from our experience. According to our experiments fitting Weibull distributions with Shiba (1969)'s data based on Birnbaum's model, the above recommendation seems to be reasonable.

If we want to have a test which seeks to be measurable only for the higher scores, the distribution with lower value $m$ of the Weibull one to the direction for higher abilities will be desirable, and it may be composed of lower difficulty and moderate discriminating power of test. And, if we want to have a test which seeks to be measurable both for the higher and lower scores, and if we want to divide the subjects into two parts, a mixed Weibull distribution would be a suitable choice.

### Table 1
Comparison of observed distribution with three theoretical distributions

<table>
<thead>
<tr>
<th>Raw score $x$</th>
<th>Estimated score on wpp $x$</th>
<th>Observed distribution</th>
<th>Kashiwagi†</th>
<th>Lord</th>
<th>Keats</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weibull</td>
<td>Beta function series</td>
<td>Neg. hypergeometric</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>85</td>
<td>102.9</td>
<td>93.0</td>
<td>133.8</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>114</td>
<td>111.7</td>
<td>104.5</td>
<td>108.5</td>
</tr>
<tr>
<td>28</td>
<td>3</td>
<td>112</td>
<td>108.8</td>
<td>106.5</td>
<td>94.0</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>103</td>
<td>100.8</td>
<td>101.4</td>
<td>82.9</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>85</td>
<td>90.4</td>
<td>92.2</td>
<td>73.7</td>
</tr>
<tr>
<td>25</td>
<td>6</td>
<td>78</td>
<td>79.2</td>
<td>80.9</td>
<td>65.8</td>
</tr>
<tr>
<td>24</td>
<td>7</td>
<td>74</td>
<td>68.1</td>
<td>69.0</td>
<td>58.8</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
<td>50</td>
<td>57.7</td>
<td>57.6</td>
<td>52.4</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>40</td>
<td>48.3</td>
<td>47.2</td>
<td>46.7</td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>48</td>
<td>40.0</td>
<td>38.3</td>
<td>41.5</td>
</tr>
<tr>
<td>20</td>
<td>11</td>
<td>37</td>
<td>32.8</td>
<td>31.0</td>
<td>36.8</td>
</tr>
<tr>
<td>19</td>
<td>12</td>
<td>27</td>
<td>26.6</td>
<td>25.2</td>
<td>32.4</td>
</tr>
<tr>
<td>18</td>
<td>13</td>
<td>22</td>
<td>21.5</td>
<td>20.8</td>
<td>28.5</td>
</tr>
<tr>
<td>17</td>
<td>14</td>
<td>16</td>
<td>17.2</td>
<td>17.5</td>
<td>24.9</td>
</tr>
<tr>
<td>16</td>
<td>15</td>
<td>17</td>
<td>13.7</td>
<td>15.3</td>
<td>21.6</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>14</td>
<td>10.8</td>
<td>13.7</td>
<td>18.6</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>10</td>
<td>8.5</td>
<td>12.5</td>
<td>15.9</td>
</tr>
<tr>
<td>13</td>
<td>18</td>
<td>12</td>
<td>6.6</td>
<td>11.6</td>
<td>13.4</td>
</tr>
<tr>
<td>12</td>
<td>19</td>
<td>10</td>
<td>5.1</td>
<td>10.9</td>
<td>11.3</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>14</td>
<td>4.0</td>
<td>10.0</td>
<td>9.3</td>
</tr>
<tr>
<td>10</td>
<td>21</td>
<td>5</td>
<td>3.1</td>
<td>9.1</td>
<td>7.6</td>
</tr>
<tr>
<td>9</td>
<td>22</td>
<td>4</td>
<td>2.3</td>
<td>8.0</td>
<td>6.1</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>3</td>
<td>1.8</td>
<td>6.9</td>
<td>4.8</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>2</td>
<td>1.4</td>
<td>5.6</td>
<td>3.6</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>7</td>
<td>1.0</td>
<td>4.3</td>
<td>2.7</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>6</td>
<td>0.8</td>
<td>3.1</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>4</td>
<td>0.6</td>
<td>2.0</td>
<td>1.3</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>2</td>
<td>0.4</td>
<td>1.2</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>1</td>
<td>0.3</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0</td>
<td>31</td>
<td>0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

† The values of $\hat{m}$ and $\hat{a}$ are 1.30 and 11.59 respectively.
with lower and larger $m$s will be desirable (this may be a rectangular type distribution). Thus, if we have a definite purpose of testing, we will be able to prepare the distribution for it.

If the subjects are planned to be divided into two parts at the preassigned ratio, the following design of the frequency distribution of mental test scores could be helpful. Suppose that we have a pool of sub-tests which are defined to be undimensional from the factor analytical point of view, and that each value $m$, the shape parameter of the Weibull distribution, is estimated. For example, if we want to divide the subjects into two parts at the level of about lower twenty-five percent, it will be recommended to construct a test in which the three-fourth subtests have about $m = 3.0$ and the rest one-fourth have about $m = 1.0$. And it will be desirable that the range of test scores in each test should be almost the same. Of course, if we assume to make use of the informations by Lord's or Birnbaum's simulation model for mental test scores, the procedure above for the design of the frequency distribution based on items of test may be applied.

**NUMERICAL EXAMPLES**

In this section, we will present three examples which show the usefulness of the Weibull distribution for the analyses of mental test scores. In the first example, Ex. 1, the Weibull distribution show a beautiful fit with the W.M.I. (cf. Keats, 1964). In the second example, Ex. 2, we will apply this distribution to Lord (1952)'s data and will show that it is possible to
define the difficulty of test by the value of the shape parameter $m$. Shiba (1969) made various kinds of distributions of mental test scores by using Birnbaum’s simulation model (cf. Lord & Novick, 1968). In order to gain the empirical relationships between the distributions of test scores and the value of the shape parameter of the Weibull distribution, and in order to make the interpretation for the value of $m$ easy, our method will be applied to his data. The results are shown in Ex. 3.


We present our analysis for W.M.I. as the first example. The frequency distribution for the test is shown in Table 1. The wpp method was applied to $1-F(x)$. That is, we assume the distribution to be measurable for the lower scores. The result is shown in Fig. 1-a. The values of $\hat{m}$, $\hat{a}$, and the smallest value of $\hat{x}$ on the wpp were 1.30, 11.59, and 1.00 respectively. Then
The Weibull Distribution and a Practical Analysis of Mental Test Scores

63

TABLE 3
Shiba's experimental distributions based on birnbaum's logistic Model

<table>
<thead>
<tr>
<th>Test Score</th>
<th>Test A</th>
<th>Test A-a</th>
<th>Test B</th>
<th>Test C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f(x)%</td>
<td>F(x)%</td>
<td>f(x)%</td>
<td>F(x)%</td>
</tr>
<tr>
<td>0</td>
<td>.8</td>
<td>.8</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>1</td>
<td>2.6</td>
<td>3.4</td>
<td>8.7</td>
<td>12.2</td>
</tr>
<tr>
<td>2</td>
<td>5.2</td>
<td>8.6</td>
<td>13.3</td>
<td>25.6</td>
</tr>
<tr>
<td>3</td>
<td>8.4</td>
<td>17.0</td>
<td>16.0</td>
<td>41.6</td>
</tr>
<tr>
<td>4</td>
<td>11.5</td>
<td>28.6</td>
<td>16.1</td>
<td>57.6</td>
</tr>
<tr>
<td>5</td>
<td>13.9</td>
<td>42.5</td>
<td>14.0</td>
<td>71.7</td>
</tr>
<tr>
<td>6</td>
<td>14.8</td>
<td>57.2</td>
<td>10.9</td>
<td>82.5</td>
</tr>
<tr>
<td>7</td>
<td>13.9</td>
<td>71.1</td>
<td>7.6</td>
<td>90.1</td>
</tr>
<tr>
<td>8</td>
<td>11.5</td>
<td>82.6</td>
<td>4.8</td>
<td>94.9</td>
</tr>
<tr>
<td>9</td>
<td>8.5</td>
<td>91.0</td>
<td>2.7</td>
<td>97.6</td>
</tr>
<tr>
<td>10</td>
<td>5.2</td>
<td>96.3</td>
<td>1.4</td>
<td>99.0</td>
</tr>
<tr>
<td>11</td>
<td>2.6</td>
<td>98.8</td>
<td>0.5</td>
<td>99.5</td>
</tr>
<tr>
<td>12</td>
<td>0.8</td>
<td>100.0</td>
<td>0.1</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Weibull's
Shape parmt. (m) 2.9 2.4† 1.9 (mixed at lower 65%)

Birnbaum's
Item parmt. (a) 0.4 0.4 0.7 1.0
Item parmt. (b) 0.0 1.0 0.0 0.0

† The value m of the Weibull distribution for the lower scores is estimated to be over 4.0.

\[ f(\hat{x}) = \frac{1.30}{11.59} \hat{x}^{0.30} \exp(-\frac{\hat{x}^{1.30}}{11.59}) \]

was obtained. Here the values of \( \hat{x} \) are integers instead of continuous values. If we substitute \( \hat{x} \) into \( 31-x \), we will estimate the distribution from the original test scores. Both in Table 1 and Fig. 1-b, we present the non-cumulative distribution for comparison with the observed one. It seems to give an impressive fit (compare with the fits by Lord and by Keats in Table 1).

Ex. 2. An application to Lord (1952)'s data: a definition of difficulty of test.

We applied the Weibull distribution to three kinds of data by Lord. They were difficult (Test 2), median difficult (Test 5), and, easy (Test 8) tests. The distributions and several kinds of statistics are presented in Table 2. It should be noted that the values of the distributions were estimated theoretically by Lord's model. They suggested that the magnitude of \( m \) was simply related to the difficulty of test (cf. Fig. 2).

Ex. 3. A psychometrical interpretation of the value of \( m \): Weibull fits with the experimental data by Shiba.

Shiba\(^2\) performed the experiments concerning the distributions of mental test scores using Birnbaum (cf. Lord & Novick, 1968)'s latent trait model. Birnbaum assumed that the distribution was mainly determined by the discriminating power and difficulty of the test. When \( a \) and \( b \) are defined as the quantities related to the discriminating power and difficulty of the test each, and when they are given appropriately, various kinds of distributions may be produced. Here note that his model is well known as the logistic one for test scores.

Our illustrative examples are Test A, Test A-a, Test B, and, Test C. The non-cumulative, cumulative-percent distribu-

\(^2\) I acknowledge Dr. Shiba for providing me his experimental data.
Test A: \( a = 0.4, b = 0.0 \)

Test B: \( a = 0.7, b = 0.0 \)

Test A: \( a = 0.4, b = 1.0 \)

Test C: \( a = 1.0, b = 0.0 \)

FIG. 3—a. Shiba's Experimental Distributions based on Birnbaum's Logistic Model.

FIG. 3—b. Shiba's Experimental Distributions based on Birnbaum's Logistic Model.

The item parameters by Birnbaum's model, and other statistics including the value of the shape parameter \( m \) of the Weibull distribution are presented in Table 3. And, for easy understanding, the non-cumulative distributions are shown in Fig. 3—a. The Weibull fits with them are shown in Fig. 3—b. These examples suggested that the value of \( m \) is simply related to the discriminability of test. That is, the greater the value of item parameter \( a \) was, the smaller the value of \( m \) was, when the value of item parameter \( b \) was fixed at zero. See Table 3 and Fig. 3—b. This may correspond to the fact that we have already pointed out in the previous section by using Lord's illustrative data. When the value of \( a \) was over .70, the distribution became a mixed Weibull one and was slightly J shaped. And when it reached to 1.00, the mixed Weibull one was more strongly J shaped at the level of lower fifty percent than Test B the value \( a \) of which is .70.

Even if a test shows rather low discriminability, and if an appropriate item parameter \( b \) is given, the test may be a measurable one to an assumed direction of measuring trait. When we tried to fit Test A—a with a simple Weibull distribution for the higher scores, the value of \( m \)
became about 2.4. This value was rather small compared to the one of Test A, i.e., about 2.9. However, when we tried to make a Weibull fit for the lower scores, the value of \( m \) was estimated to be over 4.0. See Fig. 3–b. Then, if we have a purpose of ordering subjects for the higher scores, it may be true that such a kind of test is recommendable.

**References**


Shiba, S. 1969 The distributions of mental test scores by Birnbaum’s model (unpublished material).


(Received June 9, 1969)