ワイブル分布関数によるテスト得点分布の解析：

信頼性工学における故障関数の理論と方法の導入*

日本国有鉄道労働科学研究所

柏 木 繁 男**

序

ふるくは保険数学に端を発したといわれる信頼性工学の故障関数に関する研究がさかんに行われるようになった*** (三臂; 1960; Barlow *et al*, 1967)。この故障関数に関する1つの見通しのよい数学的基礎は、Davis (1952) によってあたえられたという (三臂, 1960; McGill *et al*; 1965 Barlow *et al*; 1967)。ここでは、Davis による故障関数の理論が、基本的な改訂なしにテスト得点分布の解析に適用できることを紹介する****。しかも彼の考えが、Lord による第1種フレドホルム型積分方程式 (近藤; 1963) モデルによる潜在特性の理論 (Lord; 1965) と対照的な微分方程式モデルによる直接的接近法であることをしめす。

現在の故障関数に関する研究の発展は、長い応用面での研究と経験に支えられた扱いやすい有効な故障関数群が準備されたことがあざかっている。とくに、信頼性工学からの要請もあって、それほどの専門的技術や知識なしでも故障関数によるシステムの信頼性の管理が可能になったという。また、Weibull (1951) の故障関数による管理は、現在常識化している電子計算機の使用を必ずしも前提としないワイブル確率紙による図表解が主で、しかも、学問的研究や応用研究に十分にたえうるもので

ある。この恩恵がテスト得点分布の解析に適用できることをしめす。

テストの条件付得点率と観測された得点の分布

説明のために、正答数か正答率 (たとえば、解答関を考慮した修正正答率でも勿論よい) が問題となるテストの観測得点分布をとりあげてみよう。また、テスト得点の下限は0に無限に近づくことはあっても0とならず、つねに正の得点をとるものとする。またさらに、1ケのアイテムまたは単位となるテストを問題とするときは、能力分布に対するそのアイテムへの反応率が同時分布としてとらえられているならば十分であるが、普通このことは望めない。そこで、ここでは、テストは、とくに断わらないかぎり、いくつかのアイテムまたは単位テストの合成されたものを考えることにする。そのようなテストの各アイテムまたは単位テストがいくつか選択肢をもち、偶然選択の確率を仮定せねばならないときは (Lord, 1959), 下限点は0より大きくなることは普通だから問題はない。もし、下限点がどうしても0になるときは、適当な変換によって、上述の下限点に対する要請をみたすようにすればよい。テストの内容は、時間制限による認知図形的なものでも、学力検査的なものでもかまわない。これで、得点 x の上限点が決定されれば、 x の定義域は決定される。ここでは故障関数の理論を導入するために、上限は ∞ をとることにする。

さて、通過残存関数 $R(x)$ の定義からはじめる。

$$\left. \begin{aligned} R(x) &= 1 - F(x) \\ R(0) &= 1 \end{aligned} \right\} \quad (1)$$

$R(x)$ をていねいにかくと

$$R(x) = 1 - F(x) = 1 - \int_0^x f(t) dt = \int_x^{\infty} f(t) dt$$

となる。ただし、 $f(x)$ は $F(x)$ の x に関する導関数で、 x と t を入れ換えた表現となっている (三臂; 1960, 真壁; 1966)。また、 $F(t)$ は得点 t までの合計人員比で累加百分比でもある。したがって、 $R(t)$ は得点 t 以上

* Recommendation for use of the "Weibull distribution for the analysis of the frequency distribution of test scores: an application of the theory of the failure distribution of system reliability.

** By Shigeo Kashiwagi (Labour Science Research Institute, Japan National Railroads)

*** この研究は、三臂 武 (国鉄・技研・計算センター) との共同になるものである。彼との討論なしには成立しなかったことをつげくわえておく。

**** McGill *et al* (1965) は、反応時間の解析に故障率の概念を導入しているが、テスト得点の解析にこの概念が導入された例については知らない。

をとった合計人員比であり、 $R(x)$ は単調減少型関数となる。これで $R(t)$ が得点 t 以上の通過残存人員比、すなわち、 t における通過残存人員比をしめすことが了解されたと思うし、 $R(x)$ を通過残存関数とよんだ理由が了解されたであろう。

つぎに条件付得点率をあらわす関数を定義しよう。

$$\lambda(x) = \frac{F'(x)}{1-F(x)} = \frac{f(x)}{1-F(x)} = \frac{-R'(x)}{R(x)} \quad (2)$$

これから

$$\lambda(t) = \frac{-R'(t)}{R(t)}$$

がえられる。ここで、 $F(x)$ 、 $R(x)$ 、 $f(x)$ 、それに、 $\lambda(x)$ はいずれも微分可能な関数で、 $F'(x)$ 及び $R'(x)$ は $F(x)$ 及び $R(x)$ の x に関する導関数をあらわす。式(2)はテスト得点の増加に対する条件付得点率をしめし、 $\lambda(t)$ は得点 t におけるそれで、テストの困難度をしめす1つの有効な指標となる。すなわち、テスト得点 t と得点 $t+\Delta t$ との間で得点するであろう人員数の、得点 t 以上をとる人員数との比を、増分得点 Δt でわった量をしめす。すなわち、 t と $t+\Delta t$ の極限された Δt による面積 $[R(t+\Delta t)-R(t)]/R(t)$ を Δt でわった量である (McGill, 1965; Barlow *et al*, 1967)。

式(2)によって、テスト得点分布に対するわれわれの基本的態度がしめされた。すなわち、式(2)から $R(x)$ は微分方程式の解としてつぎのようにあらわされる。すなわち、

$$\left. \begin{aligned} R(x) &= \exp\left\{-\int_0^x \lambda(t) dt\right\} \\ F(x) &= 1 - \exp\left\{-\int_0^x \lambda(t) dt\right\} \end{aligned} \right\} \quad (3)$$

がみちびかれる (Davis, 1952; 三鶯, 1960)。このように、テストの得点分布は、式(2)の条件付得点率 $\lambda(x)$ に対する数学的仮定によってはっきり定義されることがわかる、すなわち、 $\lambda(x)$ がきまれば、 $R(x)$ 、 $F(x)$ 、それに $f(x)$ が自動的にきまる。これで、われわれの立場が $\lambda(x)$ に関する微分方程式をとくことを出発点としているという意味で、のちにふれる Lord の積分方程式による接近法と対照的であるということがしめされた。

式(3)は保険数学や信頼性工学であつかう故障関数の数学的定義をあたえるものである。(三鶯; 1960, McGill *et al*; 1965, 真壁; 1966, Barlow *et al*, 1967)。正答数のみが問題となるテストでは、 $F(x)$ 、 $R(x)$ 、 $f(x)$ 、それに $\lambda(x)$ が一般に離散型でとりあつかわれる。保険数学や信頼性工学でも、Davis (1952) 以前は離散型と連続型がちゃんぽんに論じられており、彼によって連続型への数学的拡張がなされたという (三鶯;

1960, McGill *et al*; 1965)。勿論、離散型の場合も式(1)~(3)式までの関係はそのまま通用するから (三鶯; 1960, Barlow *et al*; 1967)、以後、Davis の立場に従うものとする。

テストの得点が、仮定される能力に対して正の相関関係にあり、しかも分布が正規分布にちかく、極端に正または負に歪んでないものとしよう。逆に、課題遂行に要した時間で能力を評価しなくてはならないときは、得点は時間に対して逆の相関関係にあるものとする。もし、あるテストが、ある集団にとってやや困難なものならば、 $\lambda(x)$ は得点の増加に対して急上昇し、以後にぶるという経過をとるであろう。これに対して、テストがやや易しいものならば、 $\lambda(x)$ は得点の増加に対してはじめひくく、以後かなり急上昇するという経過をとるであろう。この傾向は、テストの得点分布が正規型に対して正または負にひどく歪まないかぎり経験に矛盾しないといえよう。これに対して、分布が正に極端に歪んで分布するような場合、すなわち、ある集団にとって極端にむずかしいテストを考えてみよう。そのときの $\lambda(x)$ は得点の増加に対して増加傾向をとるとはかぎらない。また、かなりむずかしいテストでは $\lambda(x)$ に変化があらわれず、どの得点レベルでも $\lambda(x)$ は x 軸に対して平行に動く場合も考えられる。さらに、極端にやさしいテストでは、極端にむずかしいテストとちょうど逆の傾向をしめすであろう。

得点の平均・標準偏差、それに、平均通過率などによってテスト・アイテムの困難度を考えるよりも、 $\lambda(x)$ のテスト得点の変化に対する動向は、われわれの直観力にうつつえるよりもすぐれた性質をもっている。たとえば、テスト得点の平均・標準偏差がひとしくとも、 $\lambda(x)$ のことなることもありうるのである (三鶯; 1960)。

テスト得点分布と混合・

複合型分布関数モデル*

前節では、テスト得点分布が単一な理論分布関数によって説明される場合について論じたが、ここでは、2つの関数によって表現されることが仮定されるテストの得点分布について論ずることにする。

テスト得点分布は、2つの関数、 $F_1(x)$ 及び $F_2(x)$ の混合型であり、それぞれの関数の得点分布にしめる割合を p_1 及び p_2 とするならば、その得点分布は

$$F(x) = p_1 F_1(x) + p_2 F_2(x) = \sum_{i=1}^2 p_i F_i(x) \quad (4)$$

* この節の説明は真壁 (1966) に負うている。

となるであろう。ただし、 $p_1 + p_2 = \sum_{i=1}^2 p_i = 1$ とする。

この式(4)を以後混合型分布となづけることとする。この式(4)の条件付得点率は

$$\lambda(x) = \frac{f(x)}{1 - F(x)} = \frac{\sum_{i=1}^2 p_i f_i(x)}{1 - \left[\sum_{i=1}^2 p_i F_i(x) \right]} \quad (5)$$

となる。この混合型に対してつぎのような分布が考えられる。すなわち、密度関数が

$$f(x) \begin{cases} f_1(x); \delta_1 > x \geq \gamma_1 \\ f_2(x); x \geq \delta_1 (> \gamma_2) \end{cases} \quad (6)$$

であたえられる。ただし、 δ_1 は

$$F_1(\delta_1) = \int_{\gamma_1}^{\delta_1} f_1(x) dx = F_2(\delta_2) = \int_{\gamma_2}^{\delta_1} f_2(x) dx$$

のようにあたえられる分離のパラメーターである。また、条件付得点率は

$$\lambda(x) = \begin{cases} \frac{f_1(x)}{1 - F_1(x)}; \delta_1 > x \geq \gamma_1 \\ \frac{f_2(x)}{1 - F_2(x)}; x \geq \delta_1 (> \gamma_2) \end{cases} \quad (7)$$

となる。この種の分布は混合型得点分布と区別する意味で、複合型得点分布とよぶことにする。なお、分布を2分割から n 分割に拡張することは理論的に可能であろうが、実用的見地からするとその有効度には限度があるので、ここでは、 n 分割への拡張については論じない。

テストの観測得点分布とワイブル分布関数

テストの得点分布にどういう理論的得点分布を考えるべきか、という問題はふるくから論じられてきた。たとえば、心理学者の正規分布に対する執着は異常なまでに強固であるというができるかもしれない。それはさておき、テストの得点分布をいかなる理論分布と考えるべきかは、実証すべき問題というよりは仮定すべき問題というべきであろう。この点 Lord (1965) の、観測得点分布から、真の得点や潜在特性としての能力分布を、本来分離して推定すべきであるという思想は注目しよ。Lord の上述の思想に対して彼の能力分布の推定の実際はどのようなであろうか？ この点の理解が筆者の得点分布解析への基本的態度に関連すると思われるので、簡単な説明をこころみよう。

Lord (1965) の提案は Keats & Lord (1962) の研究経験にもとづいている。彼らは、能力と得点が線型回帰関係にあり、能力を固定したときの得点分布 $h(x/\zeta)$ に二項分布を仮定し、さらに、能力の分布 $g(\zeta)$ に2パラメーターの不完備型ベーター分布を仮定すると、観測得点

分布は負の超幾何分布をすることをしめした。すなわち、観測得点分布の密度関数

$$\begin{aligned} f(x) &= \int_0^1 g(\zeta) h(x/\zeta) d\zeta \\ &= \int_0^1 \zeta^{\alpha-1} (1-\zeta)^{\beta-1} \left[\binom{n}{x} \zeta^x (1-\zeta)^{n-x} \right] d\zeta \\ &= \binom{n}{x} \frac{1}{B(\alpha, \beta)} B(\alpha+x, \beta+n-x) \quad (8) \end{aligned}$$

をみちびいた。ただし、 $x=0, 1, \dots, n$ 、それに $B(\alpha, \beta)$ はパラメーター α 及び β によるベーター分布をしめす。さて、能力密度関数 $g(\zeta)$ のパラメーター α 及び β は観測得点分布の第1及び第2次モメント u_1' 及び u_2' から

$$\alpha = \frac{u_1'(u_2' - u_1')}{u_1' - u_2'} \quad \text{と} \quad \beta = \frac{1 - u_1'}{u_1'} \alpha$$

のように推定される (Shiba & Matsushima; 1965, Lord; 1965)。この α 及び β によって連続する ζ の値に対して能力分布 $g(\zeta)$ がもとめられる。したがって、式(8)から観測得点分布への当てはめが完了する。

以上であきらかなように、式(8)の最初の右辺によって観測得点分布が、能力分布と固定した能力での得点分布による第一種フレドホルム型積分方程式で表現されていても、精密な近似解をうることは一般に困難なことが多いので (Lord, 1965)、実際には、不完備型ベーター分布による能力分布の推定から観測得点分布への当てはめをすることになる。同様のことは、Lord (1965) の4パラメーターの不完備型ベーター分布モデルについていうことができる。彼は式(8)の拡張として、能力を固定したときの得点分布 $h(x/\zeta)$ に複合二項分布を仮定し、さらに、能力分布 $g(\zeta)$ に4パラメーターの不完備型ベーター分布を仮定した。すなわち、

$$\begin{aligned} f(x) &= \int_a^b \frac{(-a+\zeta)^{\alpha-1} (b-\zeta)^{\beta-1}}{A^{\alpha+\beta-1} B(\alpha, \beta)} \left[\binom{n}{x} \zeta^x (1-\zeta)^{n-x} \right. \\ &\quad \left. + k\zeta(1-\zeta)c_2(x) \right] d\zeta; (x=0, 1, \dots, n) \quad (9) \end{aligned}$$

を仮定した。ここであきらかなように、能力分布の密度関数は、 $0 \leq a < b < 1$ 、 $\alpha > 0$ 、 $\beta > 0$ の4パラメーターによる不完備型ベーター分布である。なお、 A は $b-a$ を k は観測値 x と能力値 ζ の相関値をあらわす推定パラメーターであり、 c_2 は複合二項分布によりきまる係数である。なお、重要なことは、式(9)では能力分布の推定が4パラメーターでされることになっていても、Lord (1965) によると、 $a=0$ 及び $b=1$ 、すなわち、 α 及び β のみの不完備型ベーター分布以外の推定法では、特別の場合をのぞいて実際の解がえられないし、さらに多パラ

メーターのベーター分布を必要とするようなときは、むしろ別種の理論分布を採用すべきことを示唆している。

Keats & Lord, それに, Lord の能力分布が観測得点分布の制限をうけることは、これでほぼ了解されるであろう*。したがって、本来の目的の能力分布と観測得点分布の分離には心ずしも成功しているとは言いがたく、むしろ観測得点分布そのものの1つの当てはめ法の提案にほかならないのではないかということになる。このように、能力分布と観測得点分布の分離が必ずしも成功しているとは言いがたいのではないかということから、筆者はふるくから論じられてきた直接に観測得点分布を解析するという考えを、ワイブル分布関数の適用という観点から見なおしてみることにした。

すでに序でも指摘したのであるが、この接近法は、式(3)にみられる微分方程式モデルによる観測得点分布の解析である。このさい再び強調しておきたいことは、Lord の積分方程式モデルを解くことによりえられる解は、よい近似解がつねにえられる保証がなく (Lord; 1959, 1965), その積分方程式表示は単なる形式的記述にとどまっているといって過言ではない。これに対してわれわれの微分方程式モデルは、ワイブル分布関数を仮定した、解可能の条件付得点率 $\lambda(x)$ を使用しようという具体的提案であるという点である。また、これも序でのべたことであるが、ワイブル分布がいろいろに現象する観測得点分布に幅広い適用が可能であることは、信頼性工学の経験から当然と思われるし (Barlow *et al*, 1967), 素人にも解析しやすい形で普及されはじめてるので、テストの得点分布の解析への準備はすべて整っているといつてよいであろう。

信頼性工学における故障関数には、いろいろな形状分布が論じられてきたという、二項分布、指数分布、ガンマー分布、ワイブル分布、トランケートされた正規分布、修正極値分布、それに、対数正規分布など (Barlow *et al*; 1967)。このうち、対数正規分布の $\lambda(x)$ は作動時間に対して必ずしも単調に増加しないので、特別な性質の故障にかざられるという。故障の性質によってとりあげられる分布がことなるべきであるというのは当然であるが、故障の性質は必ずしも明白でないので、適用すべき分布に何をとりかか問題となる。そのようなとき

* Lord (1965, p. 252) は、観測得点分布形をちょっと見ただけでは能力分布形を予想することは必ずしもできないことにふれている。しかしこれは、当てずっぽうの比率、すなわち、極端な観測得点に対する制限の与え方で、能力分布はかなり自由に変動するから Lord; 1965, p. 249), 能力分布が観測得点分布と本質的にことなることの根拠とはなりえない。

もっとも有効なのがワイブル分布であるという。

テスト得点分布はいろいろな形状となつてあらわれる。あるいは正規型に、あるいはポアソン型に、テストの内容や集団の性質によりことなつた分布となる。しかし、テストの内容や集団の性質の予知が必ずしもうまくいかず、それらの性質をうまく説明することが困難なことが多い。このようなテストがワイブル分布によって解析すべきことは、信頼性工学の経験からして当然ではないかと思われる。

さて、ワイブルによる通過残存関数は

$$R(x) = \exp\left\{-\frac{(x-r)^m}{\alpha}\right\}, \quad x \geq r \\ = 1, \quad x \leq r \quad (10)$$

である。上式で $(x-r)/\alpha^{1/m} = X$ とおきなおせば

$$R(X) = \exp\{-X^m\}, \quad X \geq 0 \\ = 1, \quad X \leq 0 \quad (11)$$

となる。式(11)からあきらかなように、ワイブル分布は指数分布の拡張と考えることができる (Gumbel; 1958)。たとえば、式(10)において、 $\alpha=1$, $m=1$ のときは指数関数そのものである。 m は分布の形状をきめるパラメーター、 α は尺度のパラメーター、それに r は原点からのずれをしめす位置のパラメーターである。なお、 $\alpha=16$, $m=4$, それに $r=0$ のときのワイブル分布は正規分布にもっとも近い形状をしめす1例となるという。また、ガンマー分布はワイブル分布によく似ているが、 $\lambda(x)$ は同じではない (真壁; 1966)。ワイブル分布では

$$\lambda(x) = \frac{m}{\alpha} (x-r)^{m-1}, \quad x \geq r; 0, \quad x \leq r \quad (12)$$

となる。式(12)はテスト得点分布の条件付得点率をもっともうまく記述するものであらうと想像される。すなわち、テストの困難度や弁別度に応じて、 m の変化に応じて、条件付得点率 $\lambda(x)$ を説明したときに論じたような傾向を、はなはだ忠実にしめしうからである (真壁; 1966, Barlow *et al*; 1967)。そのほか、ワイブル分布関数による平均・標準偏差など、その他の統計的なことがらについては、たとえば、真壁 (1966) を参照されたい。

テスト得点分布のワイブル確率紙による解析*

ワイブル分布関数による得点分布の解析による効用のすべてを論ずることはできないが、いままでにわかった代表的話題を中心に、とくに、ワイブル確率紙による解

* 国鉄労働科学研究所の統計委員会テスト分科会での討議と、各メンバーの作業協力に負っている。

析（以後、ワイブル解析とよぶ）が簡便に行えて、しかも有効性がはなはだかいことを紹介しよう。

前節ですでに簡単にふれたように、ワイブル分布関数を決定するパラメーターは、形状、尺度、それに位置をしめす m , α , そして r の3つである。したがって、正確には、これら3つのパラメーターについてくわしい値を知ることが要求される。とくに、混合あるいは複合型のワイブル分布関数の場合は、単一ワイブル分布関数に比して、尺度パラメーター α のちがいが最終結果にあたる影響は無視できない場合がありうるといわれている（真壁，1966）。しかしながら、テスト得点の分析では、得点の尺度値そのものよりは、相対的な得点値が問題とされることが多い。ここで例示するまでなく、ふるくから標準得点による尺度値を無視した相関値分析が、他の科学分野では想像もつかないほどに、テスト関係の理論や研究につかわれてきたことからしても、上述のことが不合理でないことがわかる。経験的にも、尺度パラメーターと位置パラメーターの最終結果にあたる影響は、これからのべるテスト得点分布の解析において、それほど大きいものでないことが知られている。しかも、混合及び複合型の場合でも、この2つのパラメーターの影響は予想よりも小さいものであることもたしかめられている。

テスト得点分布の解析では、その分布の歪みや尖り具合に関する情報は、尺度値や原点の位置以上に重要視されることが多い。従来は、平均、標準偏差、歪度、それに、尖度などが計算さればらばらに参考にされたが、それらが統一的にあつかわれる測度をみつけることが困難であったといえよう。これに対して、ワイブル分布関数の形状パラメーター m は、これらの情報を統一した測度として、テスト得点分布の解析に有力な道具となることがわかってきたのである。しかも、ワイブル確率紙による得点分布の解析こそが最尤法やモーメント法に期待できない精度と便利さを持ち、とくに、形状パラメーター m の決定にもっとも有効であることがたしかめられているのである（Horvath; 1968）。

さて、ワイブル解析に必要なワイブル関数について簡単に情報をまとめておこう。前節の式 (11) から

$$\ln \ln \frac{1}{R(X)} = m \ln X \quad (13)$$

がみちびかれる。式 (13) の m は、 $\ln - \ln \ln$ 平面上の直線となることをしめすから、テスト得点分布が単一ワイブル分布であらわせるならその得点分布は m によってしめされる一直線上にならぶ。また、 m は分布の形をきめるから、 m の増加にともなって、正に歪んだ分布から分

散の小さい正規分布へというふうに変形する。なお、 m が1より小さくなれば勾配の急な単調減少型分布となるがこのことはすでに指摘した。

いままでの情報をもとにして、テストの困難度を、単一ワイブル分布関数の条件付得点率 $\lambda(x)$ 、すなわち、この分布の形状パラメーター m で比較することの可能性について検討しよう。最初の資料として Lord (1952, p. 45) のものをとりあげる。

Lord は、テストの得点が0点から最高8点から11点までのもので、通過率がそれぞれ .09, .34, .43, .68, .71, .90 の3つのテスト、Test 2, 5, それに、8を、それぞれ順に“むずかしい” (difficult), “中程度” (medium), そして“やさしい” (easy) と分類した。この3つのテストの観測得点から、Lord は、彼の理論による当てはめ分布を算出した。この当てはめ分布を Table 1 にしめたが、この分布に対して、それぞれ原点を移動して単一分布としてワイブル解析をしたのが Fig. 1 である。この結果によると、Test 2, 5, それに8の形状パラメーター m はそれぞれ 1.7, 3.6, それに 4.0 をあきらかに大きくこえた値となり、 $F(x)$ だけで、ごく手軽にテストの困難度の考察ができることがしめされた。

つぎに、どのような m が仮定される能力の測定に有効なのかを考察してみよう。すなわち、あるテストが測定の道具として有効かどうかを、ワイブル分布関数の m で検討できるかどうかを考えてみよう。実は、信頼性工学の経験では、 m があまりにも大きい値をしめすならば、対象となっている集団へのテストとして不向きであると考えうるのである。このことがテストの場合にも通用とするならば、そのような大きな m の値をしめすテストは、測定の道具として有効でないと解釈した方がよいことになろう。しかし、このような考えはかなり経験的で直観的なものということもできるので、さらに、テストのための理論から m の評価がされるならば、より説得力のあるものとなるであろう。

心理テストが測定の道具として有効であるということ、を、以後、測定可能という概念で説明しよう。この測定可能という概念は、テストの困難度や測定方向とは必ずしも同一のものではない。しかも、これら3つの概念はお互に重なっていて、完全に独立のものでもない。また、心理テストを施行する以上、何らかの意味で潜在する集団の能力を仮定しているのであるから、その能力の分布と実際のテスト得点の分布の形状に、なんらかの関連が仮定されてしかるべきかもしれない。

Lord (1952) は、前節の式 (8) の積分方程式モデル

を、さらに一般化したときの、誤差を無視した真の得点分布を

$$f_i(x) = \int_{-\infty}^{\infty} g(\zeta) h(x/\zeta) d\zeta \quad (14)$$

であらわすとき、能力分布 $g(\zeta)$ と $f_i(x)$ との相関度の変化によって、 $f_i(x)$ の分布がどのように変化するかにふれた。すなわち、 $g(\zeta)$ が正規分布で、しかも、 $g(\zeta)$ と $f_i(x)$ との累加正規型回帰が完全で、 $h(x/\zeta)$ はその回帰線上にうまく集中するというような理想的な形を考えたとき、テスト・アイテム i での $g(\zeta)$ と $f_i(x)$ との相関を R_i であらわすならば

$$f_i(x) = \frac{B}{\sum_{i=1}^n R_i A_i} \quad (15)$$

となるというのである。なお、 n はアイテム数、 A_i はテストの困難度によって、また B は $g(\zeta)$ の分布に対する仮定によってきめられるものである。さらに彼は、すべてのアイテムの困難度を .50 とし、また、すべての R_i をひとしくしたとき、それを、.20, .45, .65, そ

して、.80 と変化したときの $f_i(x)$ を例示した (Lord; 1952, p. 38)。これによると、 $g(\zeta)$ と $f_i(x)$ の相関が小のときは、尖度のたかい正規分布となるが、相関が増大するにつれて尖度が次第に小となり、相関が .50 をこすと、 $f_i(x)$ は“かまぼこ”型に変形して正規分布からはずれ、ついに U 字型分布になるとした。この関係をワイブル解析すると、形状パラメーター m は、 $g(\zeta)$ と $f_i(x)$ の相関の増大とともに減少する。この関係は、あるテストが、特定の集団テストとして測定可能かどうかを決定するにはなほだ示唆的なものである。すなわち、 m のあまりにも大きいテストは、ねらっている能力の測定にはむかないという有力な根拠を提供していることになる。

実用的には、因子分析法によって、共通因子空間へのテストの寄与の大きさで、測定可能の範囲をきめる指標とすることができる。たとえば、Table 1 にしめした Lord (1952) のデーターのワイブル解析によると、共通性の小さい因子妥当性の低い (low validity) Test L

Table 1
Comparison of the Theoretical Frequency Distributions of Test Scores (Lord : 1952)

Test Score	Test 2 (Difficult)		Test 5 (Medium Difficult)		Test 8 (Easy)		Test h (High Validity)		Test L (Low Validity)	
	N	Cum.	N	Cum.	N	Cum.	N	Cum.	N	Cum.
0	38.1	28.3%	4.5	3.3%			7.1	5.3%		
1	36.9	55.5	11.2	11.6			17.3	18.0	0.7	0.5%
2	25.2	74.1	16.6	23.9	0.3	0.2%	20.5	33.2	3.9	3.4
3	16.0	86.0	20.0	38.7	0.8	0.8	20.3	48.2	11.0	11.5
4	10.1	93.5	20.9	54.1	2.2	2.4	18.5	61.9	21.5	27.5
5	5.7	97.7	20.5	69.3	4.1	5.5	15.8	73.6	30.9	50.3
6	2.9	99.7	18.8	83.2	7.6	11.1	12.9	83.2	32.5	74.4
7	1.0	99.9	14.8	94.1	12.9	20.6	9.9	90.6	23.3	91.6
8	0.1	100.0	8.0	100.0	20.1	35.5	6.9	96.0	10.2	99.1
9					27.9	56.1	4.5	88.9	1.8	100.0
10					33.0	80.5	2.0	100.0		
11					26.4	100.0				
Total	136		136		136		136		136	
Mean	1.70		4.21		8.85		3.93		5.43	
S. D.	1.65		2.13		1.83		2.46		1.59	
Skewness	1.0		0.0		-1.0		0.4		-0.2	
Kurtosis	0.6		-0.9		0.7		-0.6		-0.2	
Weibull's										
Shape m	1.7		3.6		Over 4.0*		2.1		4.0	
Loca. r	-1.5		-3.0		-4.0		-1.3		-1.0	

*) The value of m for the reversed scale is 1.5.

Fig. 1

Comparison of the Theoretical Frequency Distributions of Test Scores (Lord : 1952)

の形状パラメーター $m \approx 4.0$ は、共通性の大きい因子妥当性のたかい (high validity) Test h の $m \approx 2.1$ よりもあきらかに大きいことは、Fig. 1 にしめたとおりである。この例でもそうであるが、経験的にみても、測定可能な m の上限は 4.0 前後ではないかと想像される。

測定可能の概念について、ここでは、つぎのことを留意すべきことをつけくわえておきたい。すなわち、ワイブル分布関数の形状パラメーター m で測定可能な範囲をきめるときは、困難度の検討の場合とことなり、測定方

向に無関係の m をとるようにすべきである。なぜなら、ワイブル解析では、尖度のたかい正規分布と負に歪んだ指数型や増加分布関数は、ときとして区別することができない。そのようなときは測定方向を逆にしたときの m が有効となる。たとえば、Fig. 1 における測定方向を逆にした“やさしい” (easy) テスト Test 8 は、測定方向を逆にしてはじめて測定可能なことがあきらかになったものである。

いままでのことは、正答数や正答率が問題とされる、いわば、正答型テストともいべきテスト得点分布のワ

Table 2
Comparison of the Frequency Distributions of PF_1 : Accident Groups*

pf_i	Reference Group '60		Accident Group '67		pf_i	Engine Driver (Tokyo) '65			
	N	Cum.	N	Cum.		Non-Accident		Accident	
-3.0	27	4.1%	2	5.6%	0			1	5.6%
					1				
5.0	115	21.5	1	8.3	2	3	15.0%	3	22.2
					3	5	40.0	2	33.4
7.0	192	50.5	9	33.3	4	3	55.0	4	55.6
					5	2	65.0	3	72.2
9.0	150	73.2	5	47.2	6	2	75.0	2	83.3
					7	3	90.0		
11.0	86	86.2	6	63.9	8			1	88.9
					9	1	95.0	1	94.5
13.0	41	92.4	4	75.0	10				
					11				
15.0	22	95.8	2	80.6	12				
					13			1	100.0
17.0	14	97.9	5	94.4	14				
					15				
19.0	7	98.9			16				
					17				
21.0	1	99.3			18	1	100.0		
					19				
21.1 -	6	100.0	2	100.0	20				
					21				
Total	661		36			20		18	
Mean	7.71		10.09			5.65		5.17	
S. D.	4.17		4.90			2.59		2.76	
Weibull's									
m	1.5		2.4			1.0		3.3	
r	3.0		0.0			1.6		-3.0	

*₁) Provided by Japan National Railroads.

イブル解析であった。しかしワイブル解析の効力は、誤答数、誤答率、偏差値などの解析にその本領がいかになく発揮されるものである。ただし、困難度の比較では、正答型ワイブル解析の場合と逆を考慮しておく必要がある。すなわち、ワイブル分布関数の形状パラメーター m が大きいほど困難度がたかくむずかしいテストとなることをしめすから、正答型の場合とちょうど逆となる。結局、正答型テストで能力を時間で評価するときと同様、誤答型反応と仮定される能力の間には逆相関関係にあると考えておけばよい。

誤答型テストにおいて、誤答反応の多いもの程劣るという片側棄却の考えにたつものとしよう。この考えにしたがえば、誤答型テストにおいては、劣った集団の困難度は一般に大となるはずである。すなわち、ワイブル分布関数の形状パラメーター m は、すぐれた集団のそれより一般に大とならねばならない。上述のことを、柏木 (19

64) と Kashiwagi (1964) の内田クレペリン検査作業曲線のための測度 pf_i によるデータで説明しよう。

測度 pf_i は、柏木 (1962) と Kashiwagi (1964) による内田クレペリン検査の作業曲線に対するモデルにもとづく期待作業量からの偏差量を2乗したものを標準化したものである。これは誤答型反応分布となるが、国鉄の一般基準集団 ($N=661$, 1960) と運転責任事故集団 ($N=36$, 1967) の pf_i 分布を Table 2 にしめた。これをワイブル解析したのが Fig. 2-a である。これによると、基準集団の m は1.5で、運転責任事故集団の $m=2.4$ より小となり、後者が前者に対してかなり劣っていることをしめしている。なお、ここで使っている内田クレペリン検査は、すべて、後半15分の Ⅲ(分法である。

もう1つの Table 2 にしめた事例は、Fig. 2-a とことなり、特定の事故惹起者1人に対して、作業内容、

Fig. 2-a

 Comparison of the Frequency Distributions of PF_1 : Accident and Reference Groups

経歴などなど、統制可能な物理心理的要因が同一とみなせる対照無事故者1人をえらんで、事故集団 ($N=18$) と対照無事故集団 ($N=20$)* の比較を、国鉄の東京地区の機関士について行ったものである。この Table 2 からあきらかなように、両集団間の pf_i の平均や標準偏差ともほとんどちがわないのに、Fig. 2-b のワイブル解析では、事故集団の $m=3.3$ に対して対照無事故集団の $m=1.0$ とできることがしめされている。これで、事故集団の m がより大なり、より劣ることがしめされたことになるであろう。

以上で、困難度とワイブル分布関数の形状パラメーターとの関係を利用することにより、集団の優劣の判定も可能となることをしめた。しかしながら、誤答反応に対して、誤答は極端に少なくとも多くもいけないという両側棄却の考えを採用しなくてはならない集団、たとえ

Tokyo, Japan National Railroads, 1967
 ば、傷害による事故集団とか寛解時における精神異常集団では、上式の方法を修正して使うようにすべきであろう。そのようなとき、混合または複合型ワイブル分布で表現し、もとの分布を2つ以上の分布に分割することによって、より見通しのよい結論をうるようにできるのである。以下これについてのべよう。

測定可能な概念を使うと、テストがどういう対象集団に効力を発揮するかを検討することが可能となるのは、正答・誤答型テストともに共通のことであることをすでにあきらかにした。すなわち、この概念は測定方向に無関係に決定されるので、困難度と必ずしも関係しないことをすでにのべた (Fig. 1 を参照)。また、あるテストが有効であればあるほど、そのテスト得点分布は正規分布をくずし、単一ワイブル分布関数の m が減少することにもふれた。Table 3 の傷害による事故集団 ($N=619$, 1960) を、Table 2 の基準集団とワイブル分布関数の位置パラメーター r をそろえて打点プロットすると、Fig. 3-a の破線でしめたように $m=1.2$ となり、基準集団のよりやや小となる。これに対して、国鉄の東京地区

* 事故集団が対照無事故集団に比べて2名少ないのは、たまたま、内田・クレペリン検査によるデーターの欠損を意味している。

Fig. 2-b Comparison of the Frequency Distributions of PF_1 : Accident and Non-Accident Groups
Tokyo, Japan National Railroads, 1965

Fig. 3-a Comparison of the Frequency Distributions of PF_1 : Damage and Reference Groups
Morioka, Nagoya Japan National Railroads, 1960

Fig. 3-b

Comparison of the Frequency Distributions of PF_i : Abnormal (Remission) Group
Tokyo, Japan National Railroads, 1967, 68

の寛解時における精神異常集団 ($N=93$; 1967, 68)* の Table 3 の分布によるワイブル解析による m は, Fig. 3-b の破線にしめすように, 1.1 前後となり基準集団より小となる。これは, pf_i の極端に小さいものや大きいものしめる割合が, 基準集団のそれに比べて, より大であることをしめしている。すなわち, 基準集団の分布が, やや正に歪んだ正規分布型であるのに対して, 上述の2集団のそれは, 正に歪んだより尖度の小さい, しかも, 正負の両端がややふくれあがった“かまぼこ”型にちかい分布となっていることをしめす。このような分布に適当な分割を施さないときは, 基準集団との比較があまりはっきりしないことがある。というのは, この種の分布をひとまとめにして, 平均や標準偏差などを手掛りとして集団間の比較をしても, それら集団間のちがいはほとんどあらわれないことがあるからである。たとえば, Table 3 の傷害による事故集団をひとまとめたときの平均及び標準偏差は, それぞれ, 7.30及び4.74となり, 基準集団のそれとほとんどかわらない。また, 寛解時の精神異常集団では, 上述の傷害による事故集団

よりは, 基準集団とのちがいはよりはっきりするけれども (統計的にも差は有意であるけれども), 平均 8.82 標準偏差 5.79 となり, 大きなちがいはないかにもみえる。このような分布に対して, 仮定される能力の質は2分割される可能性があると考えて, 傷害による事故集団に対しては, 基準集団の位置パラメーター r の 3.0 から 2.3 に移動して打点プロットしたのが, Fig. 2-a の pf_i 12.0 のところから Γ 型折線となる複合型ワイブル分布関数である。これに対して, 寛解時における精神異常集団では, r を 2.6 にして打点プロットすると, pf_i 5.0 のところから Γ 型折線となる。このことは, Fig. 2-b からあきらかとなる。これらの折線はそれぞれの pf_i 12.0 で分布が2分割されることを示唆するから, 傷害による事故集団に対しては, pf_i 12.0 以下とそれ以上, 寛解時精神異常集団については, pf_i 5.0 以下とそれ以上に分割して, それぞれの平均及び標準偏差を算出したのを Table 3 にしめた。また, 基準集団に対しては, 上述の2分割に対応させて, 平均及び標準偏差を別々に算出した。それによると, pf 12.0 以下の基準集団対傷害による事故集団では, 平均及び標準偏差はそれぞれ 6.73 及び 2.33 に対して 6.23 及び 2.39 となり, 後

* 職場復帰の判定を必要とする精神病歴をもつ集団で, かなり基準集団に近いもの。

Table 3
Comparison of the Frequency Distributions of PF₁: Damage and Accident Groups*

Damage Group '60		v. s.		Reference Grope '60		Abnormal (remission) Group v. s.		Reference Group '60	
N		Cum. %		N		Cum. %		N	
pf _i	Upper Lower	Upper Lower	Upper Lower	Upper Lower	Upper Lower	Upper Lower	Upper Lower	Upper Lower	Upper Lower
1	1		0.2%						
2	6	1.2%	1.1	7		1.2%		2	
3	36	7.7	7.0	22		4.6		6	
4	65	19.4	17.5	42		11.6		5	
5	81	33.9	30.5	73		23.9		12	
6	93	50.6	45.6	102		41.1		10	
7	81	65.6	58.7	90		56.2		6	
8	68	77.4	69.6	90		71.4		11	
9	38	84.2	75.8	60		81.5		7	
10	38	91.0	81.9	47		89.4		6	
11	32	96.8	87.1	39		96.0		5	
12	18	100.0	90.0	24		100.0		4	
13			19.4%	17		25.8%		4	
14	8	32.3	93.2	9		39.4		2	
15	9	46.8	94.7	13		59.1		4	
16	8	59.7	96.0	10		74.3		-	
17	9	74.2	97.4	4		80.3		2	
18	2	77.4	97.7	2		83.3		-	
19	4	83.9	98.4	5		90.9		-	
20	4	90.3	99.0	1		92.4		1	
21	-	-	-	-		-		-	
22	2	93.6	99.4	1		94.0		-	
23	-	-	-	-		-		2	
24	-	-	-	-		-		1	
25	4	100.0	100.0	5		100.0		3	
Total	557	62	619	594	67			25	68
Mean	6.23	17.05	7.30	6.73	15.59			3.58	10.46
S.D.	2.39	7.29	4.74	2.33	4.17			1.00	6.15
Weibull's									
m			1.5	1.0	1.2			1.5	1.3
r			2.3	2.3	3.0			2.6	2.6

*) Provided by Japan National Railroads.

者の pdf_i がわずかながら小となる傾向がわかり、 pdf_i 12.0以上では、15.59及び4.17に対して17.05及び7.29となり、後者の pdf_i の異常がはっきりと目立ち、分割による効果があらわれている。いうまでもなく、寛解時における精神異常集団では上述の傾向はますます顕著となる。すなわち、基準集団対精神異常集団の pdf_i 5.0以下の平均及び標準偏差は、3.77及び0.90に対して3.58及び1.00となり、後者があきらかに小となる。また、 pdf_i 5.0以上では、8.67及び3.47に対して10.46及び6.15となり、異常傾向は傷害による事故集団に比してより極端となることがわかる。

以上、ワイブル解析の有効性をいくつかの適用例を用いてのべた。これらによって、この方法がはなはば有効性のたかいものであることが了解されたと思う。ただ、このワイブル解析のような図表解には、どうしてもある程度の任意性をみとめる必要があるが、その実用的有効性の大きさを考慮するならば、十分に補ってあまりあるものであることを指摘しておきたい。実際、最尤法やモーメント法による3つのパラメーターの推定によるワイブル分布関数による解析は、その効用の点で不十分なものであるといわれ、そういうことを考慮すると、さらに、この図表解によるワイブル解析の有効度の評価はたかいものといって過言ではないといえよう。

文 献

- Barlow, R. E. & Proschan, F. 1967 *Mathematical theory of reliability*, Wiley
 Davis, D. J. 1952 An analysis of some failure data, *J. Amer. Stat. Assoc.*, 47, 113—150
 Gumbel, F. J. 1958 *Statistics of extremes*, Columbia Univ. press.

- Horvath, W. J. 1968 A statistical model for the duration of wars and strikes, *Behav. Science*, 13, 18—28.
 柏木 繁男 1962 内田クレペリン検査の作業曲線について心研, 33, 98—100.
 Kashiwagi, S. 1964 Studies on the work-curve for the Uchida-Kaepelin test, *Percept. and motor skills*, 18, 876.
 柏木 繁男 1964 内田クレペリン検査の妥当性の客観的手法による検討, 心研, 35, 98—95.
 Keats, J. A. & Lord, F. M. 1962 A theoretical distribution for mental test scores, *Psychometrika*, 27, 59—72.
 近藤 次郎 1963 積分方程式とその応用, コロナ社.
 Lord, F. M. 1952 A theory of test scores, *Psychometric mong.*, 7.
 Lord, F. M. 1959 An approach to mental test theory, *Psychometrika*, 24, 283—302.
 Lord, F. M. 1965 A strong true-score theory with applications, *Psychometrika*, 30, 239—70.
 真壁 肇 1966 ワイブル確率紙の使い方：信頼性工学のための統計的解析 日本規格協会.
 McGill, W. J. & Gibbon, J. 1965 The general-gamma distribution and reaction time, *J. math. psych.*, 2, 1—18.
 三 觥 武 1960 故障の確率モデルについて, 国鉄能研紀要, 6.
 Shiba, S. & Matsushima, K. 1965 Beta binominal, compound binominal, and guess free models (unpublished paper).
 Weibull, W. 1951 A statistical distribution function of wide applicability, *J. appl. mech.*, 18, 293—97.

(1968年2月15日 原稿受付)

第16巻第2号正誤表

倉貫論文

頁	らん	行	誤	正
表紙	氏名らん	上から5人目	記	紀
86	右	下から行4目	いただいた。	いただいた
100	右	下から8行目	かわがらない	かわいがらない

RECOMMENDATION FOR USE OF THE "WEIBULL" DISTRIBUTION FOR THE ANALYSIS OF THE FREQUENCY DISTRIBUTION OF TEST SCORES

—An Application of the Theory of the Failure Distribution of System Reliability.—

by

Shigeo Kashiwagi

Labour Science Research Institute, Japan National Railroads

One of the purposes of this paper is to apply the theory of the failure distribution of system reliability (e. g., cf. Barlow et al; 1967) to the analysis of the frequency distribution of test scores.

Assuming that test scores, x s, correlate positively with a underlying ability, we define as follows :

$$\left. \begin{aligned} R(x) &= 1 - F(x) \\ \text{and } R(0) &= 1, \end{aligned} \right\} \quad (1)$$

where

$$F(x) = \int_0^x f(t) dt.$$

And further we define

$$\lambda(x) = \frac{F'(x)}{1-F(x)} = \frac{f(x)}{1-F(x)} = -\frac{R'(x)}{R(x)}, \quad (2)$$

which is called the ratio of success. From (2) we get

$$\left. \begin{aligned} R(x) &= \exp\left\{-\int_0^x \lambda(t) dt\right\} \\ F(x) &= 1 - \exp\left\{-\int_0^x \lambda(t) dt\right\} \end{aligned} \right\} \quad (3)$$

(Davis; 1952, McGill et al ; 1965).

Another one of the purposes is to recommend for the use of "Weibull" (1952) distribution in order to analyze the frequency distribution of test scores. The distribution has often been used in the

studies of system reliability because of its wide applicability. We define the ratio of success of this distribution as follows :

$$\lambda(x) = \frac{m}{\alpha}(x-\gamma)^{m-1}, \quad x \geq \gamma; 0, \quad x \leq \gamma. \quad (12)$$

$$\left. \begin{aligned} \text{Then, } R(x) &= \exp\left\{-\frac{(x-\gamma)^m}{\alpha}\right\} \\ F(x) &= 1 - \exp\left\{-\frac{(x-\gamma)^m}{\alpha}\right\} \end{aligned} \right\} \quad (10)$$

are obtained. Here, m , α , and γ are shape, scale, and location parameters. The shape parameter m of Weibull distribution plays an important role in discriminating the degree of the difficulties and validities of psychological tests. We investigate several data by Lord (1952) and it is shown that the shape parameter m is a surprisingly useful and powerful measure in psychological testing (cf. Fig. 1).

Finally, it should be noted the following. That is, our proposal which is based upon a differential equation model is comparable to Lord's (1952) integral equation model in that the former aims to analyze directly the frequency distribution of test scores and the latter to estimate true-score or latent trait distribution.

DEPENDENT BEHAVIOR IN FEMALE ADOLESCENTS : II

by

Keiko Takahashi

Kunitachi Music College

This is the second of series of reports on developmental changes in dependent behavior (d. b.).

The present study was aimed at investigating d. b. of high school girls and comparing it with that