

男子生徒の体重分布に見られる統計的性質について

國仲 寛人*

Statistical properties in weight distribution of Japanese male students

Hiroto Kuninaka*

要 旨

2007年から2013年にかけての文部科学省学校保健統計調査のデータに基づき、日本の男子生徒の体重分布の変化に見られる特徴について報告する。16歳(2012年)と17歳(2013年)の体重の累積分布を構成し、両対数表示してみると、全体的には対数正規分布の累積形に近い形になっているものの、裾の部分は直線的になることがわかった。この分布に対し、先行研究で知られている二重対数正規分布と三重対数正規分布に加え、二重パレート分布と呼ばれる分布でフィッティングを行い、AICを比較したところ、二重パレート分布が最もよく体重分布を近似できることがわかった。

キーワード：体重分布, 対数正規分布, ベキ分布, 二重パレート分布

1 はじめに

身長と体重は人間の身体的成長を特徴づける代表的な指標である。特に体重は成長だけでなく、健康状態を表す重要な指標でもある。マッキンゼー・グローバル・インスティテュートの2014年の報告によれば、世界人口のおよそ30%が肥満や過体重であるという¹⁾。肥満や過体重は、心不全や糖尿病などの病気の原因となると考えられているため、日本においても肥満の予防は重要な課題である。このことから健康診断においては、2008年以降腹囲の測定や、BMI (Body Mass Index) の算出が必須となっている。

ある集団の健康状態を統計的に把握するために、身長分布や体重分布がしばしば用いられる。一般的にこれらの分布は正規分布に従うと思われがちだが、日本の未成年者の身長や体重の分布を調べてみると、裾の長い非対称な分布になることが多い。三橋は1953年から2006年にかけてのデータに基づいて、5歳から17歳にかけての日本の子どもの身長分布と体重分布の特徴を、年齢別に調べた²⁾。その中で、身長分布は思春期以前には対数正規分布でよく近似できる一方、思春期を過ぎた後は、対数正規分布よりも正規分布の方が近似がよくなることがわかった^{2,4)}。このことは、子どもの成長期が開始する時期が、個人によってばらつきが大きいことから説明することができる⁵⁾。

一方体重分布については、新生児の体重分布が対数正規分布でよく近似できる一方で、5歳から17歳の子どもの体重分布は、複数の対数正規分布の重ね合わせで近似できることが明らかになった^{2,6)}。すな

*三重大学教育学部

わち体重分布を式で表すと、5歳から17歳の子どもの体重分布は、

$$F(x) = \sum_{i=1}^N w_i f_i(x), \quad (1)$$

$$f_i(x) = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{\log \left(\frac{x}{\mu_i} \right)}{\sqrt{2} \sigma_i} \right) \right] \quad (2)$$

という、 N 個の対数正規分布の累積分布¹の混合分布で表されることになる。ここで x は体重であり、 $f_i(x)$, w_i はそれぞれ対数正規分布と重みを表している。式(2)に含まれる $\operatorname{erf}(x)$ は誤差関数を表している。

式(1)において、 $N = 3$ の場合を「三重対数正規分布」、 $N = 2$ の場合を「二重対数正規分布」と呼ぶ。前者は台湾の土壤におけるカドミウムの空間分布⁷⁾等に見られることがわかっており、後者は都道府県単位の人口分布⁸⁾や就学児童数の分布⁹⁾等に見られることがわかっている。

三橋は2006年のデータを用いて、5歳から17歳までの体重分布を年齢別に調べた。その結果、体重分布は対数正規分布の混合分布で近似できるという結果を得た。さらに混合分布の要素となる対数正規分布の数が、13歳を境に $N = 3$ から $N = 2$ に変化することを見出した²⁾。この傾向は1986年のデータを用いても同様であったが、2006年度以降も同じような傾向が見られるかどうかは不明であった。

本論文では、三橋が用いたデータよりも新しい2007年以降のデータを用いて、男子生徒の体重データを解析し、そこに見られた興味深い結果について報告する。

2 データ解析の方法

解析した体重データは、文部科学省「学校保健統計調査」(2007年～2013年)のデータである。学校保健統計調査は文部科学省によって毎年実施されており、調査対象となるのは、ランダムに選ばれた国内の学校に属する5歳から17歳までの幼児、児童及び生徒である。例えば、2014年の調査では、7,755校の695,600人の生徒が調査対象となった。学校保健統計調査のデータは、文部科学省のホームページから入手することが可能だが、「身長年齢別分布」や「体重年齢別分布」のデータは、度数分布表の形で提供されている。例えば2014年の12歳の女子の場合、データは20kgから117kgの範囲で1kgごとの階級に区切られ、各階級に属する子どもの割合がパーミル単位で表示されている。

統計データの調べ方としては、各調査年度において年齢別にデータを調べる横断研究と、ある年に生まれた集団の示す分布を時間的に追跡する縦断研究がある。三橋の解析は、2006年度のデータを年齢別に調べたので横断研究を行ったと言える。それに対し本論文では、1996年に出生した男子生徒の、2007年(11歳)から2013年(17歳)までのデータを用いて縦断研究を行った。これは、データ解析を行った時期に得られた最新のデータが2013年のものであったことに加え、三橋の研究において、2006年生まれの子どもの縦断研究の結果が、2016年の横断研究の結果と比較しても特徴的な差が見られなかったことによるものである²⁾。

¹ これ以降、対数正規分布の累積分布のことを、単純に「対数正規分布」と書く。

3 結果

図 1 (a)には 11 歳 (2007 年) の男子生徒の累積体重分布を, 両対数表示で示した. 図中には, 式 (1) に示される混合分布 $F(x)$ だけでなく, それを構成する対数正規分布 $f_i(x)$ も実線で示した. ただし, 混合分布 $F(x)$ のどの部分がそれぞれの対数正規分布に対応するかがわかるように, 各分布を縦方向にずらして表示している. このように 11 歳の体重分布は 3 つの対数正規分布の混合分布, すなわち三重対数正規分布でよく近似できることがわかる.

また, 図 1 (b)には 15 歳 (2011 年) の体重分布を示した. この場合は 2 つの対数正規分布の混合分布, すなわち二重対数正規分布でよく近似できることがわかる. 混合分布に含まれる N の値は, 年齢が進むにつれて $N = 3$ から $N = 2$ に変化するが, このことは三橋が得た解析結果と一致している.

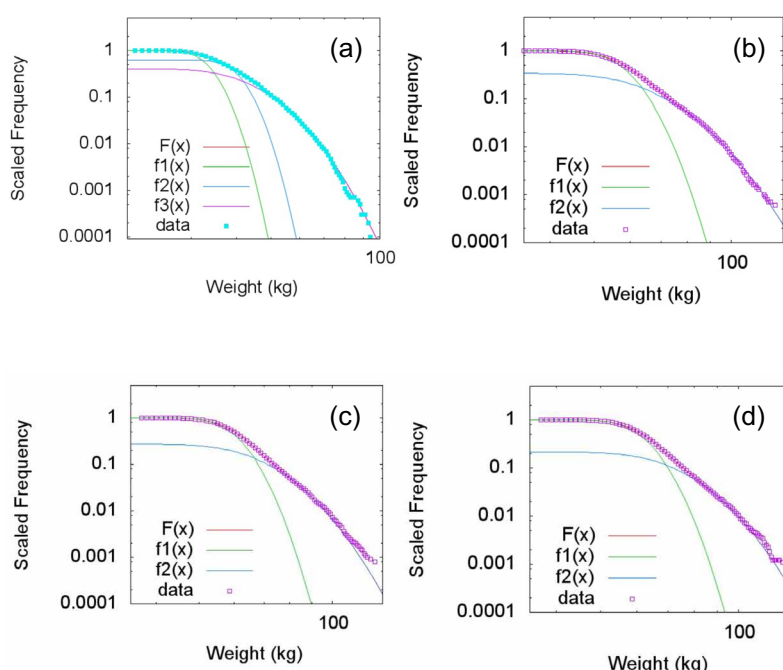


図 1 : 男子生徒の体重分布. それぞれ(a) 11 歳 (2007 年), (b) 15 歳 (2011 年), (c) 16 歳 (2012 年), (d) 17 歳 (2013 年) のデータ. 実線は, 式 (1) の分布関数およびそれを構成する対数正規分布を表している.

次に, 図 1 (c)と図 1 (d)に, それぞれ 16 歳 (2012 年) と 17 歳 (2013 年) の体重分布を示した. 各々のデータはどちらも二重対数正規分布でフィッティングを行った. ここで注意したいのは, 右裾の部分直線的になっているためフィッティング関数と分布関数の間にずれが見られるということである. つまり, この体重分布は大部分が対数正規分布で表すことができるものの, べき的な裾をもった分布関数であることがわかる. この傾向は 2013 年 (17 歳) のデータにおいても同様に見られるが, このような分布は三橋の解析では得られていない.

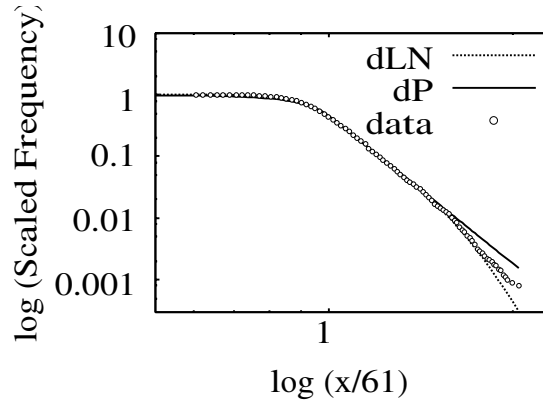


図 2 : 2012 年における 16 歳の男子生徒の累積体重分布を二つの分布関数であてはめたもの。点線と実線はそれぞれ二重対数正規分布と二重パレート分布を表している。

そこで、16 歳の男子生徒の体重分布をもう少し詳しく調べてみよう。図 2 には 16 歳の男子生徒の体重分布のみをもう一度示した。横軸の体重値は 61kg でスケールし、縦軸の頻度は総人数でスケールした。点線は二重対数正規分布を表している。この二重対数正規分布の他に、全体的に対数正規分布で近似でき、且つ裾の部分がべき的な振る舞いをする分布関数でデータをフィッティングしてみる。

対数正規分布で近似できる部分とべき的な裾を併せ持つ分布関数として、Reed らは二重パレート分布 (double Pareto distribution) という分布を提案した¹¹⁾。所得の分布¹¹⁾や、コンピュータのハードディスク内部に保存されているファイルのサイズ分布¹²⁾は、この分布関数でよく近似できることが報告されている。二重パレート分布の累積分布は、パラメータ a, b, c を用いて次のように定義される¹²⁾。

$$R(x) = \begin{cases} 1 - \frac{a}{a+b}(x/c)^b & (0 \leq x/c < 1) \\ \frac{b}{a+b}(x/c)^{-a} & (x/c \geq 1) \end{cases} \quad (3)$$

図 2 の実線は、 $a = 7.92515, b = 6.75382, c = 61\text{kg}$ というパラメータ値を用いて、式 (3) を描いたものである。この場合においても、フィッティングは全体的に良好ではあるが、やはり右裾のあたりでフィッティング関数と分布関数の間にずれが見られる。このことより、体重分布にとって最も近似がよい分布関数は見た目で判断することが難しく、何らかの指標を用いて判断する必要がある。

そこで 16 歳と 17 歳の体重分布をもっともよく近似できる分布を特定するため、複数の候補となる分布関数に対して赤池情報量基準 (AIC)¹³⁾を計算してみることにする。AIC の定義は次の通りである。

$$AIC = -2 \ln M + 2k. \quad (4)$$

式 (4) において M は最大尤度、 k はパラメータの数を表している。この値が最も小さくなるモデルが、観測データを最もよく説明できるモデルとなる。

解析の方法は次の通りである。まず、二重対数正規分布、三重対数正規分布、二重パレート分布のそれぞれで、16 歳と 17 歳の体重データをフィッティングした。そして、それぞれの最大尤度とパラメータ数に基づいて AIC を計算し、それらを比較した。

表 1 : AIC の値

Model	Age 16	Age 17
dLN (N=2)	538.2604	523.0849
tLN (N=3)	534.7753	519.1113
dP	489.0202	480.9081

表 1 に AIC の計算結果を示す. 表の中の dLN, tLN, dP はそれぞれ二重対数正規分布, 三重対数正規分布, 二重パレート分布を表している. これを見ると 16 歳の場合でも 17 歳の場合でも, 最も AIC が最小となるのは二重パレート分布であり, この分布が 16 歳と 17 歳の体重分布を表すのにふさわしい分布であることがわかる.

2007年～2013年

年齢	11	12	13	14	15	16	17
分布	tLN				dLN	dP	

2006年

年齢	11	12	13	14	15	16	17
分布	tLN		dLN		dLN		dLN

図 3 : 年齢別の体重分布の種類. tLN, dLN, dP はそれぞれ三重対数正規分布, 二重対数正規分布, 二重パレート分布である. 2006 年の表は文献[2]の結果から作成したもの.

図 3 の上段には各年齢の体重分布を最もよく近似できる分布を示した. また比較のため, 下段には文献[2]で得られた結果を示した. 12 歳, 14 歳, 16 歳の空欄は, 文献[2]で報告されていないことを示している. 年齢が進むにつれて, 三重対数正規分布から二重対数正規分布に変化するのはいずれの解析結果にも共通しているが, べき的な裾を持つ二重パレート分布は, 今回の解析ではじめて得られた分布である.

4 考察とまとめ

最後に, 体重分布にべき的な裾が出現する理由について考察する. べき的な裾の部分に含まれる体重の重い生徒の中には, 肥満の生徒だけでなく, 単純に高身長 of 生徒も含まれている可能性がある. 16 歳男子全体の約 3 %に相当する, 身長が 187cm を超える生徒の場合, 85.08kg 以上の生徒が肥満であるとみなされる¹⁴⁾. しかし, 分布関数の裾は 100kg 付近から伸びているように見えることから, この部分に含まれる大半の生徒が肥満とみなされると考えることができる.

生徒の肥満の原因は遺伝だけでなく, 生活習慣の乱れ, 運動不足など, 様々な要因が考えられる. Fildes らは 2004 年から 2014 年におけるイギリスの大規模外来データ (UK's Clinical Practice Research Datalink) から 20 歳以上の肥満の人を選びだし, その人たちが正常な体重に戻る確率が極めて低いことを見出し

た¹⁵⁾。つまり、一旦肥満と診断されると正常な体重に戻すことは困難であるということである。20歳以下の肥満の人については文献[15]では言及されていないが、おそらく年齢も近い16, 17歳の生徒の場合でも同様の傾向が見られるのではないかと推測する。

そこで、生徒の体重の変化が確率的乗算過程にしたがって変化すると仮定する。確率的乗算過程とは次のような成長過程のモデルである。

$$X_{t+1} = \alpha_t X_t. \quad (5)$$

この成長過程は、ある年の体重 X_t に、それとは独立で、ある確率分布に従う確率変数 α_t を掛けると、次の年の体重 X_{t+1} が得られることを表しており、Gibrat (ジブラ) 過程とも呼ばれる¹⁶⁾。この過程によって時間変化する量は対数正規分布にしたがうことがわかっており、幼児や児童・生徒の体重分布は対数正規分布を基本としていることから、体重の変化を表すのにふさわしい成長モデルと考えることができる。

式(5)で表される確率的乗算過程に「反射壁」を導入すると、 X_t の分布はべき分布になることがわかっている^{16,17)}。「反射壁」とは、 X_t の下限値のことである。日本の男子生徒の場合も、肥満の基準が一種の下限値の役割を果たしていると考えれば、分布の裾がべき的になることが説明できる。このようなべき的な裾の出現の仕方は、都市の人口分布がべき分布になるメカニズムに類似している¹⁸⁾。

べき的な裾の出現は、反射壁付きの確率的乗算過程で説明できるが、全体的に二重パレート分布にしたがう理由は今の所不明である。また2013年度以降のデータは未解析のため、二重パレート分布が近年に見られるものかどうかはまだわかっていない。今後、現在入手できる限りのデータを用いて、各年度の年齢別の体重分布の性質を詳細に調べてみる必要がある。

本論文では日本の男子生徒の体重分布にみられる特徴を報告した。15歳以下の体重分布は二重対数正規分布から三重対数正規分布へと変化するが、16歳と17歳の体重分布にはべき的な裾が含まれ、二重パレート分布でよく近似できることがわかった。べき的な裾は一旦肥満になった生徒が元の体重に戻りにくいという性質によるものと考えているが、その一方で1953年の女子のデータを調べてみると、6歳、8歳、11歳のデータで同様のべき的な裾が見られることがわかっている。今後は女子の体重分布も詳細に調べてみる必要がある。また、日本のデータだけでなく、肥満が深刻な問題となっているアメリカなどの海外のデータを用いて、同様の解析を行ってみるのも興味深いと思う。

謝辞

この研究の一部は、住友財団基礎科学研究助成(2012年度)を用いて行われた。

参考文献

- 1) http://www.mckinsey.com/insights/economic_studies/how_the_world_could_better_fight_obesity.
- 2) 三橋雄, 修士論文, 中央大学大学院理工学研究科 (2009).
- 3) H. Kuninaka, Y. Mitsuhashi, and M. Matsushita, *J. Phys. Soc. Jpn*, **78**, 125001 (2009).
- 4) T. Iwata, Y. Yamazaki, and H. Kuninaka, *J. Phys. Soc. Jpn*, **82**, 084803 (2013).
- 5) H. Kuninaka and M. Matsushita, *J. Phys. Soc. Jpn*, **83**, 034801 (2014).
- 6) N. Kobayashi, H. Kuninaka, J. Wakita, and M. Matsushita, *J. Phys. Soc. Jpn*, **80**, 072001 (2011).
- 7) S. Y. Yang and W. L. Chang, *Soil Science*, **170**(1), 55 (2005).
- 8) N. Kobayashi, Y. Sasaki, O. Moriyama, S. Matsushita, and M. Matsushita, *Nonlin Phen Complex Sys*, **9**(3), 276–282(2006).
- 9) H. Kuninaka, *J. Phys. Soc. Jpn*, **89**, 064801 (2020).

- 10) <http://www.mext.go.jp/english/>.
- 11) W. J. Reed and M. Jorgensen, *Comm. in Stat.*, **33**, 1733 (2004).
- 12) M. Mitzenmacher, *Internet Math.*, **1**, 305 (2004).
- 13) 赤池弘次, 甘利俊一, 北川源四郎, 樺島祥介, 下平英寿著『赤池情報量基準 AIC – モデリング・予測・知識発見 –』 (共立出版, 東京, 2009).
- 14) <http://jspe.umin.jp/medical/taikaku.html>.
- 15) A. Fildes, J. Charlton, C. Rudisill, P. Littlejohns, T. Prevost, and M. C. Gulliford, *Am. J. Pub. Health*, **105**, e54 (2015).
- 16) 林幸雄編, 大久保潤, 他著『ネットワーク科学の工具箱-つながりに隠れた現象をひもとく-』 (近代科学社, 2007年) .
- 17) M. Levy and S. Solomon, *Int. J. Mod. Phys.*, **C7**, 595 (1996).
- 18) Y. Sasaki, H. Kuninaka, N. Kobayashi, and M. Matsushita, *J. Phys. Soc. Jpn*, **76**, 074801 (2007).