

論 文

混合正規分布モデルによるマラソン完走タイムの分析
——2012年度全日本マラソンランキングに基づいて

Analysis of Marathon Time by the Gaussian Mixture Model
—— Based on the All Japan Marathon Ranking 2012

藤田尚文（高知大学教育学部）

FUJITA Naofumi

Faculty of Education, Kochi University

ABSTRACT

A Gaussian mixture model was applied to represent the marathon time distributions classified by gender and age. The model successfully predicted the distributions on the whole, but underestimated 3 hours, 4 hours, and 5 hours flat and overestimated the times such as 3 hours 15 minutes and 4 hours 15 minutes. It was assumed that runners made effort to run just within 3 hours or 4 hours. When the model was corrected in certain age ranges which caused the above discrepancies, the corrected model succeeded in predicting 96% and 87% of the data in men and women, respectively, with the precision being less than 1% error.

序

月刊誌「ランナーズ」2013年7月号付録として、2012年度全日本マラソンランキングが公表された。このランキングには男女別のタイム別順位早見表が掲載されている。年齢区分は1歳刻み、男子が20歳から90歳まで、女子は20歳から81歳まで。完走タイムは2時間45分未満から15分刻みで2時間45分以上3時間未満、3時間以上3時間15分未満と続き、6時間45分以上7時間未満までの18区分ある。例えば男子40歳の場合、完走者は7378名おり、3時間45分以上で4時間未満の者（4時間ぎりぎりの者）の順位は2574位となっている。この小冊子には早見表に続いて、男女別年齢別のそれぞれ1位から100位までの者のタイムと氏名も掲載されている。この記録はインターネットからもアクセスできるようになっている（<http://runnet.jp/runtes/ranking/>）。

本稿は、ここに掲載されている男女別年齢別のタイム別順位早見表のデータをもとに、フルマラソン参加者のマラソン完走タイムを様々な角度から分析し、その特徴を把握することをめざす。この早見表には2時間30分未満で走った人数が掲載されていないが、男子の場合、2時間30分以内でゴールした人数はかなりの数に上るので、100位までのランキング表から、年齢別にその人数を数え、完走タイム区分を19区分としてデータを分析した。

1. 男女別参加人数の分布

まず男子のフルマラソン年齢別完走者の分布をFig.1に示す。横軸は年齢、縦軸は人数を表している。全体の平均年齢は40.33歳、標準偏差は12.74歳だった。年齢ごとの完走した人数は棒グラフで表わされており、これを見ると、おおよそ正規分布に近いかたちとなっているが、20代前半あたりに小さなこぶがあるように見える。そこでこの棒グラフを単一の正規分布で近似するモデル（＝単純正規分布モデル）とふたつの正規分布を重ね合わせるモデル（＝混合正規分布モデル）で近似するモデルの比較を試みた。その結果、単純正規分布モデルの平均は40.46歳、標準偏差は12.26歳、モデルのAICは336.60だったのに対し、混合正規分布モデルは面積比0.027、平均27.15歳、標準偏差2.50歳の正規分布（赤線で表示）と、面積比0.973、平均が40.98歳、標準偏差12.23歳の正規分布（緑線で表示）のふたつを重ね合わせたモデル（＝混合正規分布モデル、紫線で表示）のAICは323.67とこちらの方が小さく、こちらの方がよいモデルであることが判明した。実際Fig.1を見るとこのモデルがほぼ実測値の分布を近似していることがわかる。46歳の棒グラフのへこみは丙午年うまれの人口減のためであろう。男子のフルマラソン完走者は「ランナーズ」誌付録にあるように21万5067人であり、このうち赤線で示された分布に含まれる人数は6千人弱、緑線で示された分布に含まれる人数は20万9千人あまりだった。

次に女子のフルマラソン年齢別完走者の分布をFig.2に示す。実測値を棒グラフで表わしているが、分布は26歳付近と39歳付近のふたつの頂点をもつ分布となっており、一瞥して正規分布をして

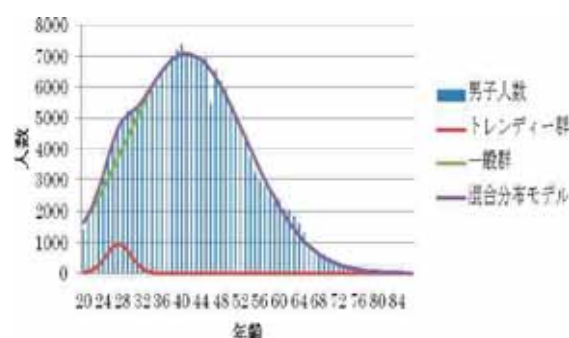


Fig.1 男子マラソン完走者の年齢分布

いないことがわかる。そこでこの分布をふたつの正規分布の重ね合わせモデル（＝混合正規分布モデル）で近似を試みた。その結果、面積比0.110、平均26.52歳、標準偏差3.42歳の正規分布（赤線で表示）と、面積比0.890、平均が39.40歳、標準偏差11.32歳の正規分布（緑線で表示）のふたつを重ね合わせたモデル（紫線で表示）がほぼ実測値の分布を近似することができた。このモデルのAICは229.76であり、単一の正規分布モデルの289.10と比べて小さかった。後者のモデルの平均は37.16歳、標準偏差は12.16歳だった。女子のフルマラソン完走者は5万8687人であり、このうち赤線で示された分布に含まれる人数は約6千5百人、緑線で示された分布に含まれる人数は5万2千人あまりだった。この前者の人数は男子における同様の分布の人数とあまり変わらない。

この男子と女子の赤線で表示された群はどのような群なのだろうか。女子の場合、20代を中心として健康とファッションを指向する者たちがフルマラソンに参入しており、この者たちが一般のフルマラソン参加者とは別の層を形成していると思われる。そこでこの20代の一部の層をトレンドィー群と名付け、これ以外の者たちを一般群と称することにする。

男子の場合もおおよそ同じで、ファッションブルな女性たちに主導されながら、必ずしもマラソンそのものを強く志向していない若者がいるように思われる。男女でこの群に属する人数があまり異なることも、このことを裏付けているかもしれない。この点については後述のマラソンタイムの分布のところでもう一度論ずることにする。

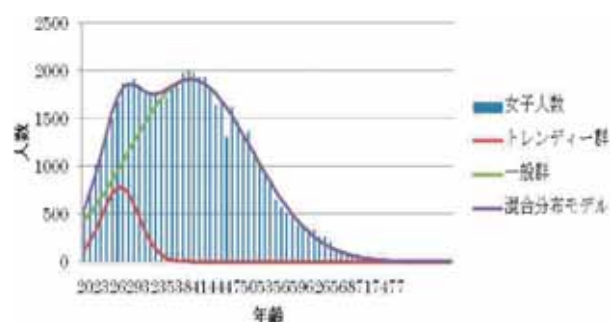


Fig.2 女子マラソン完走者の年齢分布

2. 完走タイムの分布（全体）

男子のすべてのフルマラソン完走者の完走タイムの分布を Fig.3 に示す。まず目につくのは、3 時間 45 分以上 4 時間未満のランナーの数が両隣のランナーの数と比べて格段に多いことである。完走タイムの 4 時間切り、いわゆるサブ 4 は多くの素人ランナーのひとつの目標となっており、余裕をもって走れば 4 時間少々くらいのランナーたちがおそらくサブ 4 をめざして必死にがんばるので、このバーの高さが異常に高く、結果として 4 時間以上 4 時間 15 分未満のバーの高さが低くなっているのであろう。同様のことは 5 時間付近でも起きていることが見てとれる。

次に女子のすべてのフルマラソン完走者の完走タイムの分布を Fig.4 に示す。女子の場合も男子の場合と同様、4 時間付近、5 時間付近のバーの高さが右隣よりも高く、4 時間 15 分、5 時間 15 分のバーの高さの落ち込みがあり、とりわけ 5 時間 15 分のところでこれが大きい。

6 時間から 6 時間 15 分のところにかげ、男女の両方のグラフにおいてバーの高さにかなりの隔たりがある。ここでも 6 時間切りをめざして頑張っている姿がうかがえるが、これは他方、制限時間が 6 時間の大会がかなり多数あり、6 時間以上のランナーの数が制約を受けていることが大きく影響している。

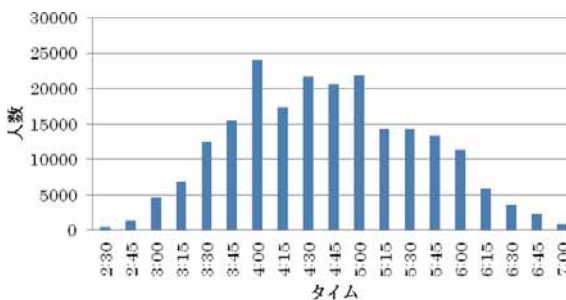


Fig.3 男子全体の分布

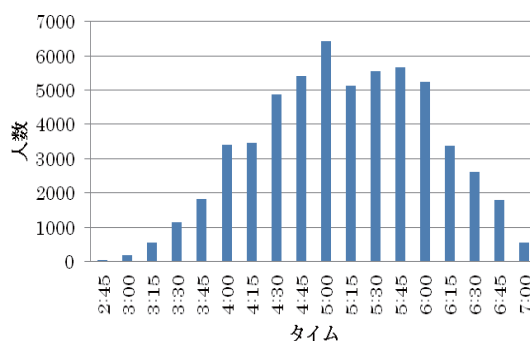


Fig.4 女子全体の分布

3. 年代別完走タイムの分布

我々が手にしているデータは、男女別年齢別のマラソン完走タイムの分布である。ここでは前節とは異なり全体ではなく、年齢ごと

に完走タイムの分布がどうなっているのかを詳しく見ていく。ただし年齢幅は 20 歳から 80 歳を超える年齢までの 60 歳以上に及ぶので、年齢を 5 歳ごとにまとめ、20 歳から 24 歳（これを 20 代前半と呼ぶ）、25 歳から 29 歳（これを 20 代後半と呼ぶ）など 70 代前半まで年齢を 11 区分にまとめて分布を調べることにする。これ以上の高年齢はデータ数が少ないのでここでは取り扱わない。これらの男女別のデータを 4 つのパネルにまとめ、Fig.5 に示す。(A)は男子の 20 代前半から 40 代前半まで、(B)は男子 40 代後半から 70 代前半まで、(C)は女子 20 代前半から 40 代前半まで、(D)は 40 代前半から 70 代前半までのグラフである。

男子においては、すでに Fig.1 でも見たように、年齢が 20 代前半から 40 代前半にかけて完走人数が多くなっており、40 代前半で人数的なピークを迎え、40 代後半から 70 代前半まで減少し続けている。

女子においては Fig.2 で見たように、20 代後半で完走人数がいったん多くなり、そこからやや減少に転じて 30 代前半から 40 代前半にかけて再び増加傾向になり、40 代前半でピークを迎える。このことは Fig.5(C)においては、ただちには視認することができない。なぜかと言うと、男子 40 代前半までを表示する(A)では年代別のグラフがあまり重なり合わず、40 代前半で曲線が一番上に来ているので人数の増加がただちに視認できるのに対し、女子 40 代前半までを表示する(C)では 20 代前半のグラフが一番内側に来ているのを除くと、5 時間 15 分を境にグラフの左右で分布の上下関係が逆転しているという大きな特徴がある。

つまりタイムが 5 時間 15 分より速い方では 20 代後半から 40 代前半にかけて人数が増えているのに対し、5 時間 15 分より遅い方では、20 代後半の人数が一番多く、年齢が上がるにつれて人数が減少している。20 代前半と 20 代後半はこのくぼみの右側の人数の方が左側より多いのにたいして、30 代前半になるとほぼ左右の人数が等しくなり（生データでは 31 歳で左側の人数が多くなる）、30 代後半以降左の人数が多く、60 代前半以降また右側が多くなる。つまり 20 代（20 歳から 29 歳）と 60 代（60 歳から 69 歳）に 5 時間 15 分以上のタイムで走るランナーが多く、30 代前半から 60 歳未満のランナーの多くは 5 時間 15 分以内で走っているという傾向がある。また 40 代後半から 70 歳代にかけては単調に減少し続けている。

なぜ 20 代（20 代前半と 20 代後半）の女性に“遅い”ランナーが多いのだろうか。これは第 1 節で指摘した女子におけるトレンドイ群の存在と関係があるだろう。おそらくこの群の女性はランニングの練習量が必ずしも多くはなく、練習中および大会において自分の好きなファッションに身を包み、とにかく完走をめざすことを目的としているのであろう。

男子においても 20 代前半、20 代後半のランナーは 30 代前半から 60 代前半のランナーと比べて遅い傾向がある。平均タイムを調べると 30 代前半から 60 代前半のランナーは 20 代のランナーと比べて速く、60 代後半になってやっと 20 代のランナーたちのタイムと並ぶ。Fig.5(A)(B)のパネルを見ても、30 代後半から 50 代後半ま

での各折れ線グラフでピーク値は4時間であるのに対し、30代前半では4時間と5時間のところの2箇所ピークがあって5時間のピークが4時間のピークよりもやや高く、20代前半と20代後半においては5時間のところのピークが4時間のものよりもかなり高いところにある。

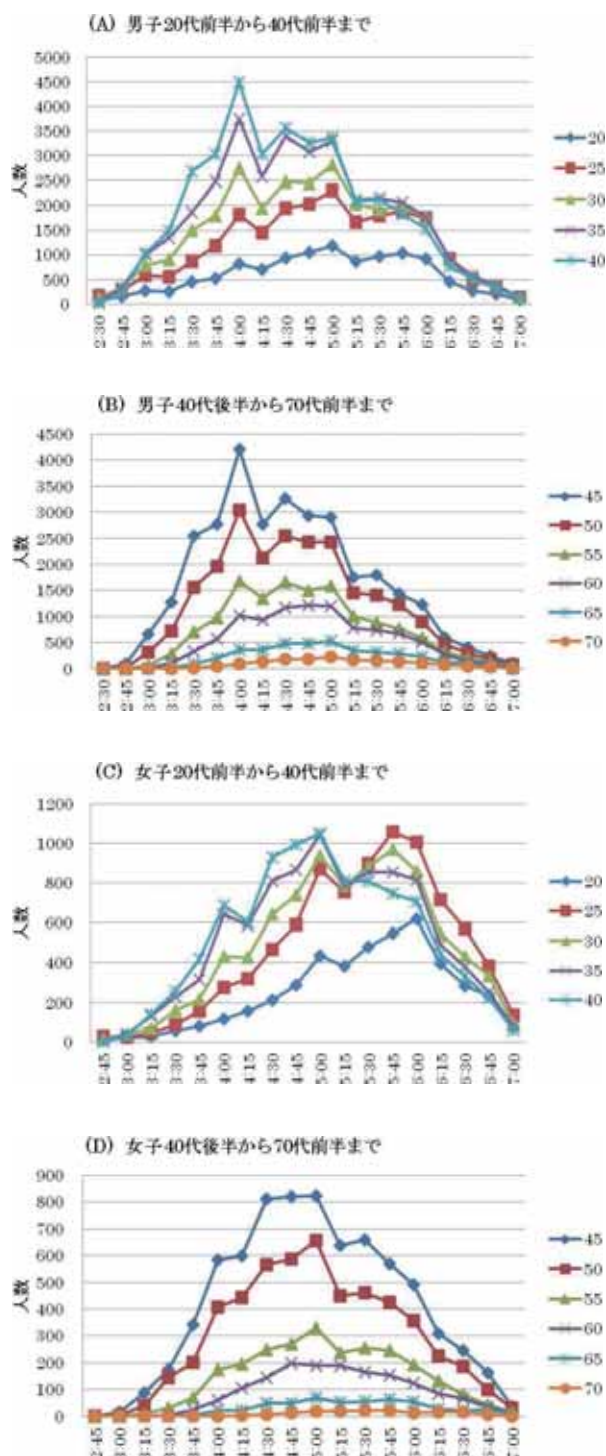


Fig.5 男女別年代別マラソン完走者のタイム分布
横軸はタイム、縦軸は人数、各折れ線は年齢を表示。

完走タイムが5時間15分以上の者は、男子20代で1万1千人、女子20代で7千4百人ほどいる。1節で指摘した20代を中心に存在するトレンドー群は男女とも6千人程度いる。Fig.1, Fig.2を見て分かるように20代のランナーのすべてがトレンドー群というわけではないが、この群が遅いランナーの主力であり、20代ランナーの完走タイムが30代以上のランナーたちよりも遅い主要な原因であると推定してよいのではないかとと思われる。(C)に典型的に見られる5時間15分の箇所のくぼみは、前節でも指摘した、サブ5をめざしてがんばる層がいるための反動として形成されたものと理解するのが適切だろう。男子においては、20代前半、20代後半にはサブ3後のくぼみ(3時間15分の箇所)が見られ、30代前半でこれが弱くなり、30代後半で消失している。しかしながら30代後半、40代前半においても3時間のところは上に凸の傾向が見てとれる。つまりサブ3をめざしてがんばる姿が少しだけ見てとれる。

男子において最大のくぼみはサブ4後のくぼみで(4時間15分)、20代前半から60代前半まで認められる。サブ4.5後のくぼみ(4時間45分)は30代前半から50代後半まで、弱いながら認められる。サブ5後のくぼみ(5時間15分)は男子においては顕著ではなく、20代前半、20代後半と40代後半に認められる。しかしながら5時間から5時間15分にかけては、折れ線がくぼんではいないものの急激な人数の減少があり、サブ5めざして頑張っているランナーが多いことが読み取れる。

女子においてはサブ4のくぼみは30代前半から50代前半まで認められる。サブ5のくぼみは20代前半から50代後半まで認められる。

6時間から6時間15分にかけて、男女ともくぼみは認められないが、折れ線グラフの傾斜が全体的に急になっており、これらが重なり合ってFig.3, Fig.4における段差となったのであろう。制限時間が6時間の大会が多くあることも、この一因である。

男女を比較すると、男子においては全体の分布がやや左側のタイムの速い方に偏っているのに対して、女子においてはやや右側のタイムの遅い方に偏っているように見える。全体として男女ともマラソン完走タイムの分布は正規分布していないように見える。またサブ4後のくぼみのような、なめらかな分布から逸脱している箇所がいくつかある。

4. 男女別年齢別マラソン完走タイプの混合正規分布モデル(第一次近似)

前節の分析を踏まえ、ここでは1歳刻みの男女別マラソン完走タイムの分布を記述できるモデルを提案する。第一次近似として、第一節と同様、ここでも二つの正規分布を重ね合わせる混合正規分布モデルを用いる。これは上で指摘したように、グラフが正規分布から逸脱していることが明瞭に見てとれることから、このモデルを採用することにした。

ここでマラソン完走タイムと月間走行距離に関する思考実験を試みよう。経験的に知られているように、月間走行距離の多いラ

ランナーには速いランナーが多く、少ないランナーには遅いランナーが多い。しかしながら走行距離の多いランナーでも相対的にあまり速くないランナーもいれば、走行距離の少ないランナーでもかなり速いランナーもいる。統計的にいえば、ある月間走行距離（たとえば 200 km）のランナーのマラソン完走タイムは正規分布をし、ある平均と標準偏差をもつであろう。そして月間走行距離の多いランナーの完走タイムは速い傾向があるだろう。そこで Table 1 のような事態を想定してみよう。

ここでは月間走行距離が 500 km のランナーのマラソン完走タイムの平均が 3.2 時間、標準偏差が 0.3 時間で、そういったランナーは全体の 1 割いと想定されている。同様に 50 km のランナーの平均タイムは 5.25 時間、標準偏差は 0.8 時間で比率が 2 割であると想定されている。

Table 1 月間走行距離とマラソン完走タイムの平均、標準偏差、比率の仮想的な例

月間 走行 距離	500	400	300	250	200	150	100	50
平均 タイム	3.2	3.5	3.75	4	4.25	4.5	4.75	5.25
標準 偏差	0.3	0.35	0.4	0.45	0.5	0.6	0.7	0.8
比率	0.1	0.1	0.1	0.1	0.1	0.15	0.15	0.2

そしてこれらの群はいずれもタイムが正規分布すると想定されている。これは仮想的な例だが、おおそ実態もこれに近いものと思われる。これらの分布が重なりあって全体の分布形が決まるので、これら 8 群と全体の分布を Fig.6 に示してみる。

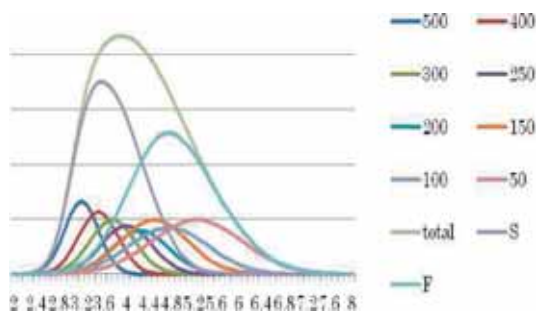


Fig.6 月間走行距離とマラソン完走タイムの仮想例
横軸は完走タイム、縦軸は比率

グラフの下の方にあるのが 8 群からなる正規分布で、一番上にあるやや左に偏った正規分布していない分布がこれらを重ね合わせた分布である。これらの 8 群のうち走行距離が 500, 400, 300,

250, 200 km の群を重ね合わせたものを S 群, 150, 100, 50 km の群を重ね合わせたものを F 群として表示している。すると本来は 8 群の重ね合わせであった全体の分布が 2 群の分布に集約され、S 群はやや左に偏っているかもしれないが、2 群はおおよそ正規分布していることが見て取れる。

ここで採用する混合正規分布モデルの基本的考え方は以上の例に尽きている。月間走行距離ごとに完走タイムの平均と標準偏差が決まるとしても、多数の群を便宜的に 2 群で表現することで、単一の正規分布による記述と比べて格段にデータを近似する力が高まるのである。複雑な形状をしている全体の分布を記述する 2 群は厳密には正規分布をしないかもしれないが、これらが正規分布していると想定しても事実上、問題がない場合が多いだろう。混合正規分布モデルは 3 群以上の正規分布を想定することも可能であるが、ここでは 2 群による表現で、どの程度実測値を近似できるかを探る。

マラソン完走タイムに影響を及ぼす要因は月間走行距離だけではない。身体的素質であったり、マラソンにかける思いであったり多様な要因があるだろう。それらは多様な群を形成しているかもしれないが、上で述べたように実質的に 2 群で表現することが可能である。それらを総称して一方の群を S 群、シリアスランナー群と呼び、他方を F 群、ファンランナー群と呼ぶことにする。

混合正規分布モデルの基本的考え方は以上だが、マラソン完走タイムの分布においては 4 時間付近や 5 時間付近で、グラフがなめらかな分布よりも明らかに高くなったり、低くなったりする傾向が存在している。この傾向はおそらく混合正規分布モデルのみでは説明できない。そこで第二次近似として、これらのデータの補正を行う。その根拠はランナーがサブ 4 やサブ 5 をめざすため、統計学的な分布からは逸脱するという心理学的な要因が存在するためである。

分析対象となるデータ

1 歳刻みの年齢区分ごと (n)、15 分刻み完走タイム (m) の完走者の人数比 $Db(n, m)$ と $Dg(n, m)$ 。これらはある完走タイムまでの完走者の人数を年齢ごとの完走者の総数で除した値。添字の b は男子、 g は女子を表す。年齢は 20 歳から 64 歳まで 1 歳刻み ($n=20, 21, 25, \dots, 64$) の 45 区分。完走タイムは男子においては 20 時間区分 ($m=2.5, 2.75, 3.0, \dots, 7.0$) (2 時間 30 分未満から 7 時間までの 15 分刻み)、女子においては 19 時間区分 (2 時間 45 分から 7 時間までの 15 分刻み)。早見表には男子において 90 歳まで、女子においては 80 歳までのデータが掲載されているが、高年齢のデータは数が少ないので、分析対象からは除いた。

混合正規分布モデル—第一次近似

2 つの正規分布の重ね合わせモデル (2 要素混合正規分布モデル) を採用した。平均値の小さい方の正規分布を g 、大きい方を G とする。ここでは累積型正規分布を用いた。分布の形を決めるパラメータは g の平均 x_n と標準偏差 s_n 、 G の平均 y_n と標準偏差 z_n 、および 2 つの正規分布の面積比 r_n と $1 - r_n$ 。

男子の場合、

$$Db^{^}(n,m)=g(m,x_n,s_n)*r_n+G(m,y_n,z_n)*(1-r_n). \quad (1)$$

で $Db(n,m)$ を近似する。 $^{^}$ は推定値を表す。女子 $Dg^{^}(n,m)$ も同様。

制約条件

隣り合う 2 つの年齢において、完走タイムの分布形は似ているはずである。言い換えると、2 つの分布 g と G のパラメータは年齢間でなめらかに変化するはずである。これを制約条件とする。

目的関数

実測値を近似する混合正規分布モデルのうち、上記の制約条件を満たす目的関数を以下のように定め、これの最小値を求めた。

男子の場合、

$$Zb = \sum \sum (Db(n,m) - Db^{^}(n,m))^2 + 0.1 * \sum \{ (x_{n+1} - x_n)^2 + (s_{n+1} - s_n)^2 + (y_{n+1} - y_n)^2 + (z_{n+1} - z_n)^2 + (r_{n+1} - r_n)^2 \}. \quad (2)$$

(m=2.5, 2.75, ..., 7)(n=20, 21, ..., 64)

女子(Zg)の場合は、 $m=2.75$ から。なめらかさに関わる後項の重みをどれくらいにすべきかは理論的には定まらないので、多少、試行錯誤したうえで、0.1 と定めた。 n は 64 までだが、データは n が 65 のときも存在するので、 $n+1$ の計算のさいにこれも利用した。

このような Zb , Zg を最小にするパラメータ x_k , s_k , y_k , z_k , r_k を求めるためにエクセルのソルバーを用いた。ひとつの最小値を求めるために、複数の初期値を試し、局所的極小解を避けるようにした。

5. 混合正規分布モデルによる近似結果

男子については 25 歳、40 歳、55 歳のモデル値と実測値のグラフ、女子については 30 歳、40 歳、50 歳のモデル値と実測値のグラフを Fig.7 に示す。モデルの値を求めるにあたっては累積型正規分布を用いて計算したが、ここでのグラフは分布型(階級幅ごとの人数比)として描かれている。棒グラフは完走タイムごとの人数比、赤の折れ線がモデルの値、黄緑の折れ線が S 群、紫の折れ線が F 群の値を示している。

混合正規分布モデルは年齢ごとの実測値のさまざまな分布形の概形を捉えているが、予想されたとおりモデルの値からの逸脱も目につく。男子の 3 つの分布形においては 4 時間、5 時間、6 時間の箇所を実測値がモデル値を上回っており、4 時間 15 分、5 時間 15 分の箇所を下回っている。また 25 歳と 40 歳のグラフで 3 時間の箇所がモデル値を上回り、55 歳のグラフで 3 時間半の箇所がモデル値を上回っている。

女子においては、3 つのグラフで、4 時間と 5 時間の箇所の実測値がモデル値を上回っており、5 時間 15 分の箇所において下回っている。また 40 歳と 50 歳で 6 時間の箇所を実測値が上回っている。ここで提示したグラフには示されていないが、女子のグラフの

上回りと下回りは男子のグラフほど安定的ではなく、たとえば 60 歳のグラフにおいては 5 時間 45 分の上回りが見られたりしている。

S 群の分布は男子の場合、最速のランナーから 5 時間 30 分くらいまでのランナーを含み、女子の場合は最速から 6 時間くらいまでのランナーを含んでいる。F 群の分布は男子の場合、一番速いランナーは 3 時間そこから一番遅いランナーまで、女子の場合は速い方が年齢によって変化しているが、3 時間台のランナーまで含む。両群はかなりの程度重なり合っており、あるランナーのタイムだけを見て、そのランナーがどちらの群に属するかを判定するのはかなり難しい。どうしてこうなるかは最初に例としてあげた、月間走行距離をイメージするとわかりやすい。150 キロ程度のランナーでも 3 時間そここのランナーがいるだろうし、200 キロでも 4 時間半程度のランナーはいるだろう。

男女の S 群と F 群の平均完走タイムが年齢とともにどう変化しているかを Fig.8 に、標準偏差の年齢変化を Fig.9 に、S 群の比率が年齢とともにどう変化しているかを Fig.10 に示す。さらに年代別平均完走タイムを Fig.11 に示す。Fig.8 を見ると、男子の S 群の平均タイムは 40 代後半まで 3 時間後半で安定的に推移しており、そこから次第に遅くなっている。F 群は、はじめ 5.2 時間くらいで 30 代から速くなり、40 代後半で一番速く、そこから悪くなっている。女子の S 群の平均タイムは 4 時間前半から 50 代前半まで少しずつ悪くなり、そこから低下の程度が強まる。F 群は年齢を通じてほぼ横ばいだが、中盤でややタイムがよくなっている。

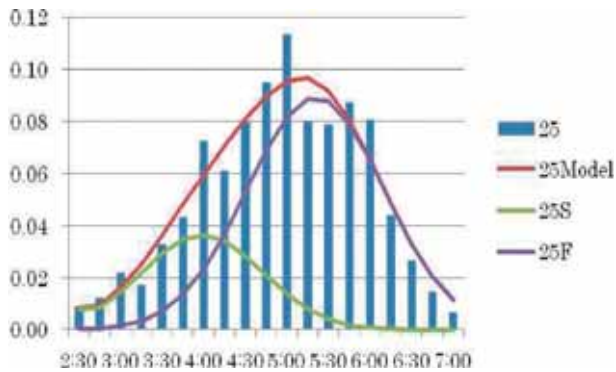
Fig.9 を見ると、標準偏差は男女とも年齢とともに小さくなっており、男子の S 群においてその傾向が強いが、40 代後半からは横ばい状態となっている。また男女とも 20 代の前半においては、値が一時的上昇している。女子の方が S 群と F 群の差は小さい。

Fig.10 を見ると、S 群比率は男女とも 20 代前半で低くなっているが、20 代半ばからは増加となっている。男子は 30 代半ばから 50 代半ばまで、ほぼ一定で 0.45 前後で推移し、50 代後半から増加している。女子は 20 代半ばで 0.1 を切るが、そこから 50 歳くらいまで増加を続け、そこから低下に転じている。

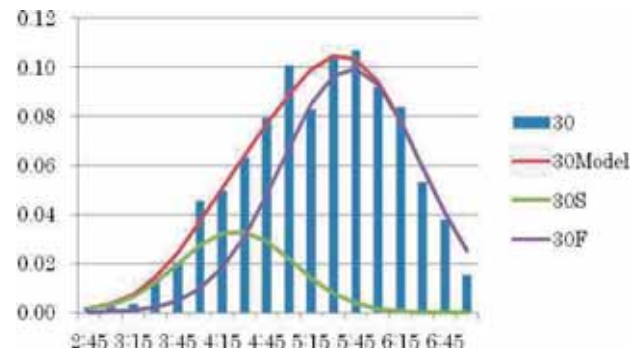
Fig.11 を見ると、完走タイムは S 群比率などと同様に 20 代前半の傾向がそれ以降と反対の向きになっており、20 代半ばからは男女とも減少傾向が続き、40 代半ばからは増加に転じている。

完走タイムと S 群比率の関係について見てみよう。両者の関係は、完走タイム = S 群比率 * S 群平均完走タイム + F 群比率 * F 群平均完走タイムである。女子の方から見ていく。S 群のグラフを上下反転するとほぼ完走タイムのグラフと重なる。しかも細かな凹凸まで一致している。S 群と F 群の平均完走タイムのグラフが総体的にはフラットなので、完走タイムの年齢による変化は、ほぼ S 群比率の変化に帰してよいだろう。男子の場合、40 代の半ばまでは、女子の場合と同様で、完走タイムの年齢による変化は主として S 群比率の変化によって説明されるが、この年齢の範囲内で F 群の平均タイムが速くなっているのも、これも一部寄与しているだろう。40 代後半以降は S 群比率の下降、S 群 F 群の平均タイムの悪化が輻輳して完走タイムが悪くなっていると理解できる。

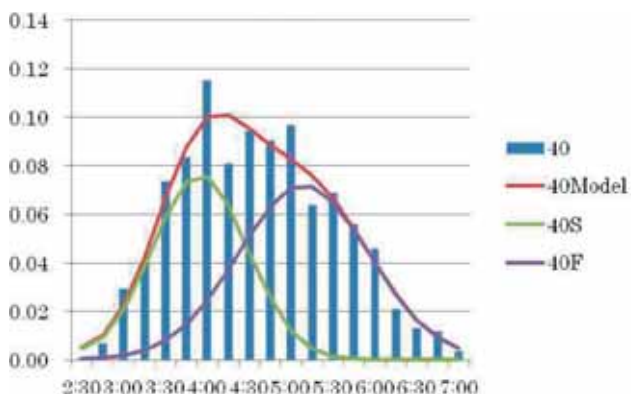
(A) 男子 25 歳



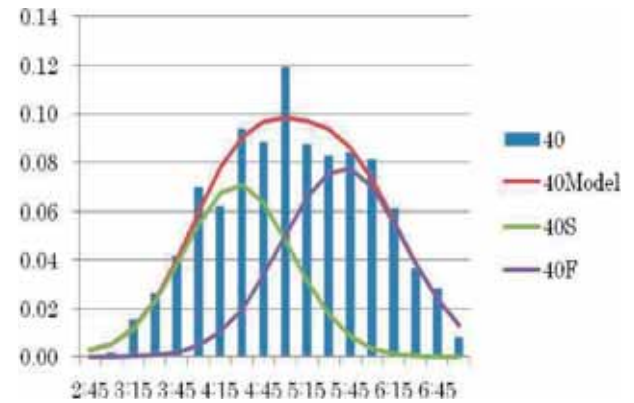
(D) 女子 30 歳



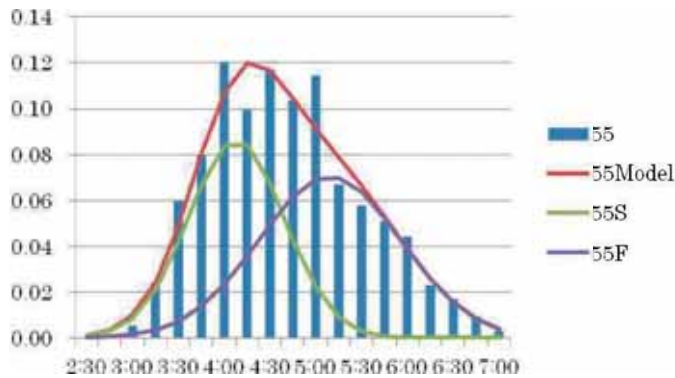
(B) 男子 40 歳



(E) 女子 40 歳



(C) 男子 55 歳



(F) 女子 50 歳

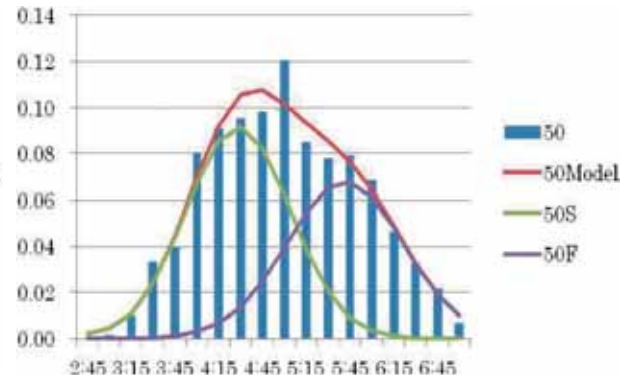
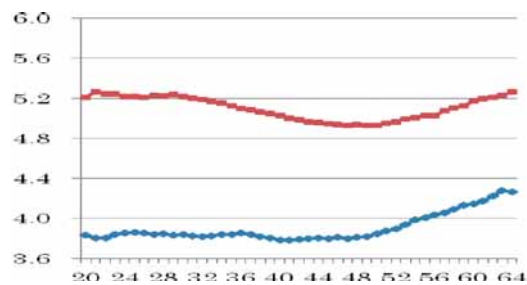


Fig.7 男女別各年齢における実測値と混合正規分布モデルの値
横軸は完走タイム、縦軸は人数比率を表す。棒グラフは実測値の比率、黄緑の曲線はS群、紫の曲線はF群、赤の曲線は混合正規分布モデルの値。

(A) 男子



(B) 女子

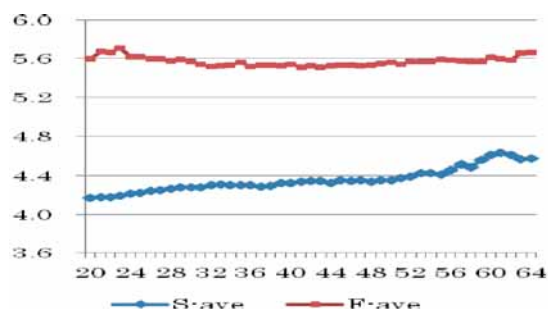
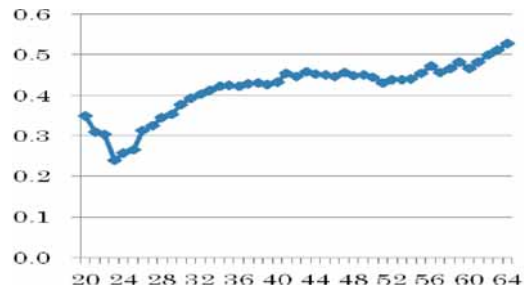


Fig.8 S群とF群の平均完走タイム

上の図(A)が男子、下の図(B)が女子。横軸は年齢、縦軸は完走タイムを表す。◆はS群の完走タイム(下の折れ線)、■はF群の完走タイム(上の折れ線)。

(A) 男子



(B) 女子

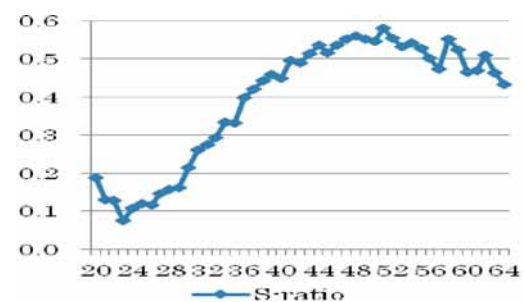
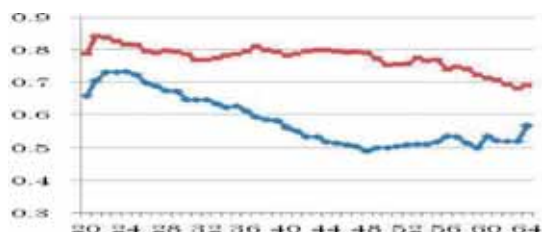


Fig.10 S群の比率

上の図(A)が男子、下の図(B)が女子。横軸は年齢、縦軸はS群の比率を表す。

(A) 男子



(B) 女子

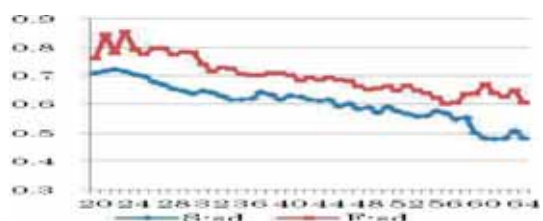
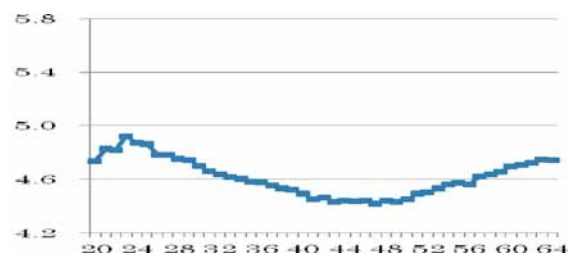


Fig.9 S群とF群の標準偏差

上の図(A)が男子、下の図(B)が女子。横軸は年齢、縦軸は標準偏差を表す。◆はS群の標準偏差(下の折れ線)、■はF群の標準偏差(上の折れ線)。

(A) 男子



(B) 女子

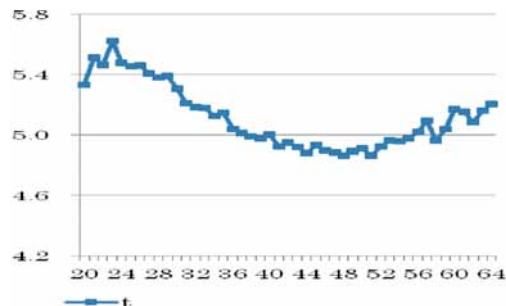


Fig.11 年代別平均完走タイム

上の図(A)が男子、下の図(B)が女子。横軸は年齢、縦軸は完走タイムを表す。

6. 混合正規分布モデルからの逸脱傾向の補正（第二次近似）

前節で述べたモデル値からの逸脱傾向は、混合正規分布モデルでは捉えられないものの、ランナー心理を考えると十分に了解可能な傾向である。そこでモデルのあてはまりを評価する際にはこれらの値を補正してモデル値を得ることとする。

補正を行う基本的方針は完走タイムのある時間区分において、ある年齢からある年齢までの範囲内の平均逸脱量を求めその値を補正值とするというものである。その間に符号が反対の箇所が含まれていても、全体として補正をすることであてはまりが改善すると判断できるときには補正を行った。このような方針で求めた補正量を Table 2 に示す。この表に示されていない時間区分においてもある年齢範囲で一定の逸脱傾向が見られる場合もあったが、ここでは補正箇所を常識的な 3, 3.5, 4, 5, 6 時間にとどめた。年齢別にあてはまりを検討すると、男子の場合、41 歳まではサブ 3 をめざしてがんばるランナーが多く、実測値がモデル値を上回る傾向が続くが、42 歳からはその傾向が消失し、代わって 46 歳からはサブ 3.5 をめざしている傾向がうかがえた。

Table 2 男女別補正量

男子			女子		
時間	年齢	補正量	時間	年齢	補正量
3 時間	21～41	0.007	4 時間	34～57	0.011
3 時間 30 分	46～56	0.010	4 時間 15 分	32～56	-0.013
4 時間	20～64	0.017	5 時間	21～59	0.013
4 時間 15 分	30～64	-0.021	5 時間 15 分	20～58	-0.012
5 時間	21～64	0.014	6 時間	21～43	0.013
5 時間 15 分	20～64	-0.014	6 時間 15 分	21～42	-0.008
6 時間	20～64	0.008			

男子において各階級において混合正規分布モデルの値に上で述べた補正量を加えた値と実測値の隔たりが±1%未満の箇所は 19 階級×45 年齢区分＝855 箇所中の 818 箇所、96%がこの範囲に収まった。女子においては上で述べたように、逸脱の傾向が男子ほど鮮明ではないために、補正を加えても±1%未満の箇所は 18 階級×45 年齢区分＝810 箇所中の 704 箇所、87%にとどまった。累積型において同様の補正を行うと（補正量は上で示した値と異なる）、あてはまりが若干改善し、男子においては 839 箇所で 98%、女子においては 736 箇所、91%が±1%未満の逸脱となった。男子の場合、ランナーズ誌のタイム別早見表と実用上遜色のない順位予想ができていたが、女子の場合、実用上は少し不満の残る精度であらう。

例として男女いくつかの補正後のモデルが実測値をどれくらいよく近似しているかを示す。Fig.12 および Fig.13 参照。

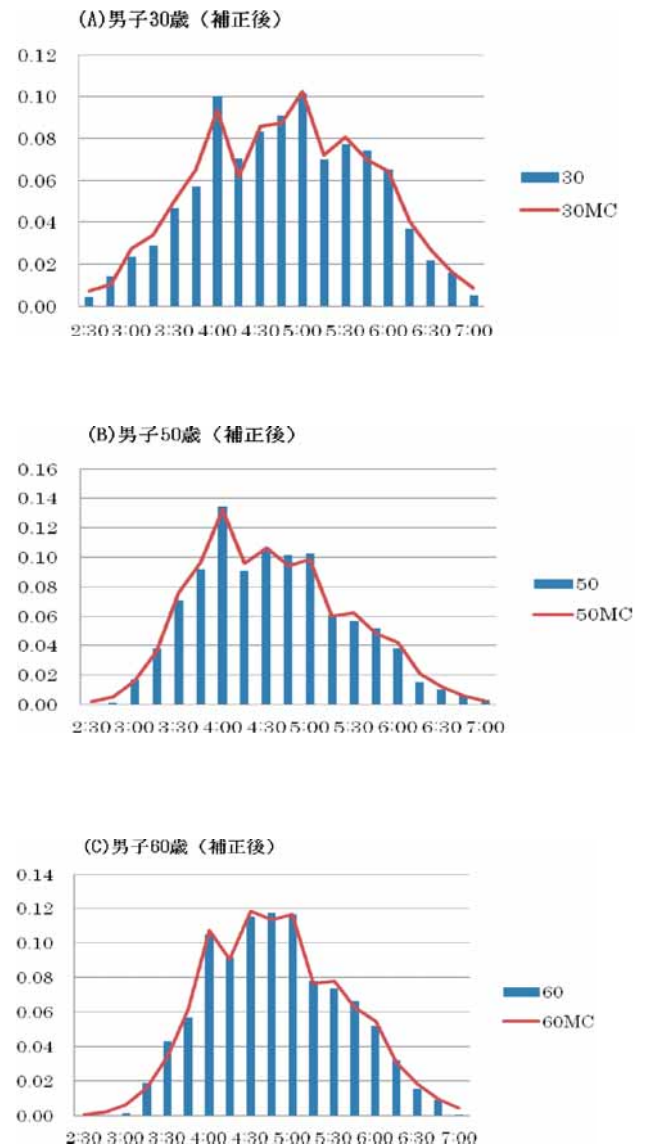


Fig.12 男女別年齢別の実測値と補正後のモデル値（男子）

7. 考察

マラソン完走タイムの分布は単純な正規分布をしないので、理論的な分析は難しかったが、混合正規分布モデルを導入し、サブ 4 などの補正をすることで、分布をモデル的に予測（近似）することができた。このことの意義について考察する。

混合正規分布モデルによる 2 群の抽出

混合正規分布モデルでデータの分布をうまく捉えることができたとしても、そのことでただちに、この分布導入の正当性が担保されたわけではない。あくまでも他のデータなどと照合して、経験的

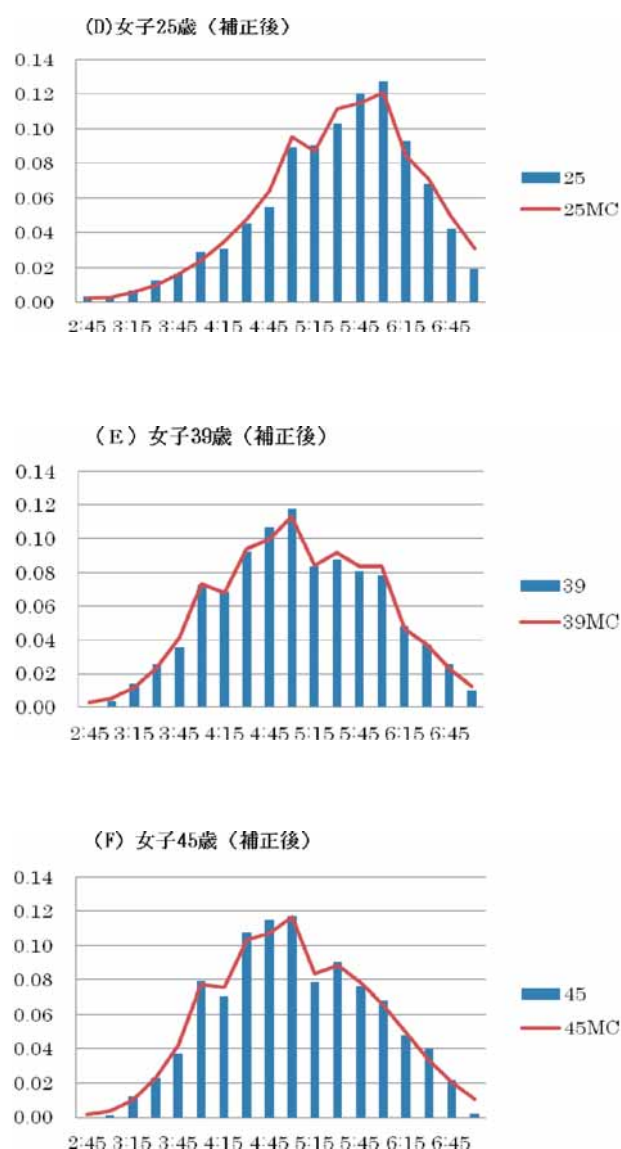


Fig.13 男女別年齢別の実測値と補正後のモデル値(女子)
横軸は完走タイム、縦軸は比率。棒グラフは実測値、折れ線は補正後のモデルの値。

に2群に対応するものがみつからなければならないだろう。世の中のマソンランナーのなかには、できるだけ速くマソンを完走したいと思い、完走タイムにこだわり、熱心に練習を重ねる人たちと、気持ちのよいジョギングをして、その延長上にフルマソンの完走を願う人たちがいることは経験的には明らかではないだろうか。前者がここでシリアスランナー群と呼ぶ人たちであり、後者がファンランナー群と呼ぶ人たちである。

この2群の存在については、単にフルマソンの完走タイムの分析だけからでは、実証できず、ランナーたちの意識調査、月間走行距離、身体能力等、多様なデータとの照合が必要となる。その意味で、ここで提案している両群は、現時点では仮説的な存在である。

サブ4補正などについて

サブ3、サブ4、サブ5、サブ6付近のデータの凸凹が存在することもランナー心理からすると当然のものと思われる。このことは、自らにふさわしい(本来的な実力の)ペースで走れば、4時間を少し超えるタイムでゴールする人で、サブ4めざしてがんばることにより、本来の実力以上のタイムで走ることができる人が多数いるということを意味する。市民ランナーレベルのランナーたちにおいては、自分の身体的な仕上がりが一緒でも、めざす目標タイムによって、かなりタイムが伸びたり縮んだりするのである。データの上では必ずしも示されていないが、逆のケースとして本来なら3時間47分で走れるランナーが少々楽をすることによって3時間58分でゴールすることもありうるだろう。

なめらかさの制約条件

本稿では混合正規分布モデルのパラメタを求めるにあたって、年齢ごとの分布に最もよくあてはまる5つのパラメタを求めるのではなく、20歳から64歳までのすべての年齢の45個の分布を見通した上で、年齢ごとに求められるパラメタがなめらかに変化するという制約条件を置き、目的関数 Z_b 、 Z_g を求めた。なぜこの条件を置いたのかという現実的な理由は、この条件を付加せずに、年齢ごとに分布のパラメタを求めると、隣り合う2つの年齢間で値がかなり大きく変化する事態が発生したためである。

混合正規分布モデルは分布の微小な変化に敏感である。このモデルを適用するデータはサブ4などの補正をする前のものであり、モデルは4時間あるいは5時間のピークに正規分布の頂点を置こうとするかもしれないし、これ以外の偶発的な増減に反応する可能性がある。あるいは月間走行距離と混合正規分布モデルの関係を例にとるなら、S群を構成する集団がいつも一定と限らず、あるときは走行距離が200km以上がS群と認定されたり、別のあるときには150km以上がS群と認定されたりする可能性がある。多数ある下位の分布を2群で代表させるとき、こういうことが起こる可能性がつかにある。

それにたいして、もしS群とF群がある程度実質のある群で例えば前者がタイムにこだわるのが集団の中核的特徴で、後者がランニングを楽しむことを特徴とする群だとしたら、2群は隣り合う2つの年齢間で大きな変化はしにくいはずで、年齢とともにゆるやかにその比率や平均タイムを変化させると想定できるだろう。ここではこういった実質的な群を追求するために、この制約条件を置くことにし、分析はうまくいったと考えられる。

年代別平均完走タイムの変化

Fig.11を見ると、40代半ばまでは平均完走タイムが向上している。これはシリアスなランナーが多数いて、そのランナーたちが日々トレーニングに励んでいるので少しずつタイムが向上すると理解すべきであろうか。Fig.8を見るとS群ランナーの平均タイムは男子はほぼ横ばいであり、女子はやや低下している。それにもかかわらず、なぜ平均完走タイムは40代半ばまで年齢とともに向上

するのか？その答えの一部はすでに述べたように、S群ランナーの比率が高くなるからである。では長年走り続けるシリアスランナーのタイムは向上していないのだろうか。

この問題を考えるために、ある年齢 n から次の年齢 $n+1$ におけるS群、F群の変化を以下のように定式化してみよう。以下の考察はS群、F群の経年変化を記述しているので、本稿で扱っている年代間の分析とは別の事象である。

$$S_{n+1} = a S_n + b F_n + c P_{n+1}, \quad (3)$$

$$F_{n+1} = d F_n + e S_n + (1-c) P_{n+1}. \quad (4)$$

S_n, F_n ・・・ある年齢 n におけるS群、F群の人数

P_{n+1} ・・・ある年齢 $n+1$ における新規参入者数

a ・・・S群であり続ける率

b ・・・F群からS群に移る率

c ・・・ P_n のうちS群となる率

d ・・・F群であり続ける率

e ・・・S群からF群に移る率

このうち、

$$a + e \leq 1, \quad (5)$$

$$b + d \leq 1. \quad (6)$$

n 歳におけるシリアスランナーが $n+1$ 歳のときに、全員ランナーであり続けるならば $a + e$ は1となる。ファンランナーにおいても同様。

$n+1$ 歳におけるS群の人数は n 歳におけるS群のうち a の割合の人が引き続きS群となり、 n 歳においてはF群だった人のうち b の割合の人がS群に移り、さらに新規参入者のうち一定割合 c の人がはじめてS群となる。このうちF群だった人は月間走行距離などが伸びたためにS群に移ったわけだが、この人たちのタイムが $n+1$ 歳のS群の平均タイムを総体として上回るとは期待しにくいだろう。同様に新規に参入した人たちも $n+1$ 歳のS群平均タイムを下回っている可能性が高い。結局引き続きS群に留まる人たちは40代半ばくらいまでは少しずつタイムが向上し、そこに新たに加わる2群の人のタイムとの平均がおおよそ一定ということが生じっていると推測される。

さらに推測を重ねると、 $n+1$ 歳のときにS群から離脱する人（マラソンをやめてしまう人とF群に移る人）はタイム的に振るわなくなった人に多いかもしれない。そうするとS群にとどまる人たちのタイムはハードなトレーニングをつまなくても平均タイムが少し向上することになり、さらに標準偏差は少し小さくなるだろう。Fig. 9の標準偏差（これはここで扱う標準偏差とは性質が異なるものだが）は年齢とともに減少する傾向にある。この人たちに上記2

群の人が加わるという可能性もある。

もしこの後者の可能性が真実なら、シリアスランナーの集団であるS群は総体としてはトレーニングの効果によるタイムの向上はしていないことになる。個々のランナーにおいては無論、努力の成果があらわれてタイムを大幅に縮めている者もいるだろうが、年齢的な衰えを自覚してタイムの低下を受け入れているランナーも多くいるということになる。シリアスランナーの中には新規参入者もいるかもしれないが、大部分は永年走り続けているランナーであり、かつ練習態度がシリアスであるために、自らの限界のタイムで走る人たちである可能性もあるだろう。

F群についてはどうなのだろうか。F群の男子は40代半ばまで平均値は向上し、女子は横ばいである。標準偏差は男子は40代半ばまであまり変化せず、女子は小さくなっている。

男子の場合、F群であり続ける人に平均タイムの向上する人が多いのではないだろうか。つまりこの人たちは、自らの限界付近でタイムを競っていないので、伸びしろはむしろS群のランナーよりも大きいのではないだろうか。この中でさらにランニングに目覚めた人がS群に移るのかもしれない。したがってF群であり続ける人の標準偏差は小さくなる傾向があるかもしれないが、新規参入者がいるために、結果として標準偏差はあまり変化しないと考えられる。

女子の場合をどう考えるか。女子は男子よりも健康志向が強く、あまりタイムに固執していないのかもしれない。女子のF群であり続ける人たちは余力を残しながら楽しく走り、トレーニング効果が出るので、あまり遅いランナーはいなくなるし、無理をしてがんばるランナーも少なくなるかもしれない。したがって標準偏差が減少するかもしれない。

なぜ年齢別上位者のタイムは若いほどよいのか

最後に素朴な問題について考えよう。なぜ年齢別上位者のタイムは若いほどよいのか？具体的に言うと、「ランナーズ」2013年7月号付録には年齢別ベスト100位までのランナーの名前が載っているが、20代のランナーのタイムと40代のランナーのタイムにはだいぶ隔たりがある。これは一見すると当たり前と思われるかもしれないが、ここまでの分析によると、男子のS群の平均タイムは20代から40代まで横ばいなのである。このことを踏まえると、この疑問は必ずしも当たり前ではないことが了解されるだろう。この答えはとりあえずS群の男子の標準偏差が若いほど大きく、年齢が上がるにつれて小さくなることに求められる。つまり若いランナーのS群標準偏差が大きいということは、とびきり速いランナーがいるということを意味する。そしてS群の平均値が変わらないまま、年齢とともに標準偏差が小さくなれば、速いランナーの数が年齢とともに減少するのである。

しかしながらこの「答え」は、答えというより、事態を統計学的に記述したにすぎないものである。本当の答えは運動生理学的説明を必要とするものであり、その答えはある意味で自明なものであるだろう。

