生と死の数理：あなたが 50 年後に生きている確率は？ [1]

半田 賢司 （佐賀大学理工学部）

数理サイエンスコースの模擬授業に出席下さり，ありがとうございます．オープンキャンパスを 3 年ぶりに対面で開催することができ，大変うれしく思います．この 2 年半ほどの間，死というものを意識させられる機会がそれ以前よりも多くあったのではないでしょうか．この講義では，生存や死亡に関する事柄を主な題材とし，その数学的な解析および応用の一端を紹介します．

高校数学との関連で言えば，データの分析や確率・統計といった内容と深くかかわるものです．特に，確率変数という考え方が重要です．それはその値が偶然性に左右されるものなのですが，ここでは，誰にとっても最も身近な確率変数を考えます．すなわち，その人が何歳で死ぬかという値，死亡時満年齢です．

利用するのは，資料としてお渡しした「厚生労働省 第 23 回生命表（完全生命表）」[2] です．完全生命表は 5 年毎に作成されていて，国勢調査による人口の結果も利用されているそうですが，この 3 月に第 23 回生命表が発表されたばかりであることは，私たちにとってタイミングが良かったと思っています．

１．生命表とは？

資料 p.8〜p.11 にある数表が生命表です．一種のモデルだと考えて良いのですが，おおまかに言うと，10 万人から成る仮想的な集団について，その生存数を 1 年毎に列挙した列が主要な部分で，他の箇所はそれから派生する数値が並んでいるものと見ることもできます．性別で分かれているという点は重要ですが，生命表の意味そのものは同じですので，説明を行う際は男の表を主に参照することにします．

p.8 の上部，左の方を見て下さい．ただし，初めの方のいくつかの行,「週」や「月」と記されている部分は見ないで「年」とある行とそれ以降を見て下さい.「生存数」「死亡数」の意味ですが，年齢 $x = 0$ で生存数 $l_0 = 100000$ であったものが，次の年，1 歳 (年齢 $x = 1$) で生存数 $l_1 = 99816$ となっています．これらから 0 歳の死亡数が $100000 - 99816 = 184$ であることが導かれますが，それはまた，右の欄に死亡数 $d_0 = 184$ とあることと同じです．一般的には，記号を用いますが，l_x は x 歳の生存数であり，$d_x = l_x - l_{x+1}$ は x 歳の死亡数，すなわち死亡時満年齢が x の人数です．引き算の意味は「x 歳になれるが $x + 1$ 歳にはなれない人の数」といったことです．

２．生命表からわかること　—「データの分析」の例として —

私たちは死亡時満年齢に注目しているのですが，この授業では，既に調べたような数値を死亡時満年齢の仮想的データとして読み直します．データの個数（サイズ）はもちろん 10 万です．つまり，10 万人の死亡時満年齢が記録されたデータです．

まず，0 歳の死亡数 $d_0 = 184$ は，全てのデータを小さい順に並べたときに最も小さい値 0 が 184 個ずらりと並ぶことを意味し，そのすぐ下の欄の $d_1 = 24$ から，次に小さい値として 1 が 24 個並ぶことになります．もしヒストグラムをイメージしたと

[1] 令和 4 年 8 月 10 日 (水) 佐賀大学オープンキャンパス 2022 における 1 時間の高校生向け模擬授業
[2] https://www.mhlw.go.jp/toukei/saikin/hw/life/23th/index.html　エクセルファイルもダウンロードできます．

したら，データ 0 に付随して 184 人分の高さを持つ棒が建ち，そのすぐ右隣にデータ 1 に付随する 24 人分の高さの棒が建ち，... といったものになります.[3]

そのような並びの中で最後に来るデータの値，言い換えると最も大きなデータの値はいくつでしょうか．p.9 の最下部を見て下さい．この部分は不思議なのですが，男の表で最後の行は $x = 113$ でその生存数は 1 であるのに対し，死亡数は 0 となっています．なお，$x = 112$ に対する死亡数は 1 です．生命表作成の詳細を調べる等してそうなっている理由を確認した訳ではありませんので，この場での取り扱いとしては，男の場合，死亡時満年齢のデータの最大値は 112 でその個数は 1，ということにしておきます.

(2-1) 重要な値を調べてみよう.

データが与えられたときにはまず重要な値について調べる，ということは高校で習っていると思います．ワークシート内の表に予めそれらを記載しておきました：

	最小値	第 1 四分位数	中央値	第 3 四分位数	最大値	最頻値	平均値
男							
女							

男のデータについて考えてみましょう.（女の場合については，後で少し時間をとりますので，各自で埋めてもらいます.）最小値と最大値は先ほど述べたように，それぞれ 0 と 112 です．ここでの**平均値**の計算は結構大変なのですが，幸いそれに相当する数値を生命表内に見つけることができます．このデータの場合，平均値は平均寿命と呼ばれるものと同じで，表の右端「平均余命」の欄で $x = 0$ のときの値です．それによると，男の死亡時満年齢の平均値は 81.56 です.

最頻値（モード）は簡単にわかります．死亡数が最も多い年齢であり，したがって，死亡数 d_x が最大となる x ということにもなりますので，表をたどって行けば $x = 88$ がそれを与えることが見て取れます．また，他に最頻値を持たないこともわかります．実際，88 歳の死亡数は 4133 であり，他のどの年齢よりも死亡数が多いことが確認できます.（後で構いませんが，d_x が極大や極小となる x がないか調べて下さい.）

今度は**中央値**（メジアン）です．これはそんなにすぐにはわからないかも知れません．データの個数が 10 万であったことを思い出しましょう．すると，中央値は 5 万人目に死ぬ人と 50001 人目に死ぬ人の死亡時満年齢の中間の値（正確には足して 2 で割った値）です．そこでこれらの人たちが何歳で死亡しているかを調べるわけですが，これは生存数が全体の半分である 5 万を切るのはいつかを見極めることと同じです．そう着眼すれば，中央値が 84 であることが容易にわかります．実際，84 歳になれる人は 51850 人いるのですが，85 歳になれる人は 48204 人ということで 5 万人を大きく切ってしまっており，これらの事実は 5 万人目に死んだ人と 50001 人目に死んだ人が同じ 84 歳で死んだことを意味しています.

[3] 実質的にこれを視覚化したものが，厚生労働省「第 23 回生命表（完全生命表）の概況」の p.5 にグラフとして見つかります．この文書の URL は次です．https://www.mhlw.go.jp/toukei/saikin/hw/life/23th/dl/23th-11.pdf

あとは**第 1 四分位数**と**第 3 四分位数**です．今度も境目を探します．第 1 四分位数については，2 万 5 千人目に死ぬ人と 25001 人目に死ぬ人の死亡時満年齢の中間の値ということですが，やはり生存数の方で見て生存数が 7 万 5 千人を切るのはいつかを調べます．すると第 1 四分位数は 75 であることがわかります．実際，75 歳になれる人は 76033 人で 76 歳になれる人は 74002 人です．同様にして，第 3 四分位数は 90 であることが，90 歳になれる人は 28082 人で 91 歳になれる人は 24096 人であることからわかります．以上の結果を基にして表を完成させると

	最小値	第 1 四分位数	中央値	第 3 四分位数	最大値	最頻値	平均値
男	0	75	84	90	112	88	81.56
女	0	83	90	95	114	93	87.71

となります．女の場合の数値については上の通りです．確かめてみて下さい．

(2-2) 箱ひげ図を描いてみよう．

箱ひげ図は，今表に書き込んだ数値の内，最頻値以外の数値を数直線上に配置して視覚化したものです．ただし，両端には髭をつけ [4]，メインボディと言える第 1 四分位数から第 3 四分位数までの箇所には箱を描きます．人体になぞらえると腰から肩までの部分でしょうか．箱内にはさらに，中央値の箇所に仕切りを設けます．また，平均値がわかっているので，その箇所に「＋」を記しておきます．一般には平均値が必ずしもこの箱内におさまるとは限らない [5] のですが，このデータの場合は箱内に位置します．

付け加えると，平均値が中央値よりも左に位置しています．[6] これから重要なことが言えます．それは，平均寿命よりも高い年齢で死ぬ人が過半数を占めるという事実です．このため，平均寿命を把握している人が新聞のお悔やみ欄を眺めたとき，そのような状況を実感する機会がしばしばあったとしても不思議ではありません．

それでは，男女それぞれについて，横向きの箱ひげ図を完成させて下さい．そんなに正確に描く必要はないのですが，男のデータの第 3 四分位数と女のデータの中央値とが一致しているということには注意を払って下さい．

(2-3) 男女の違いについて言えることは？

この箱ひげ図からデータのおおまかな傾向が見えてきます．ここでは男女のデータの比較の観点から，2 点だけ確認しておきたいと思います．まずは空欄付きで：

・男は女よりも死亡時満年齢が $\boxed{}$ 傾向にある．

・男は女よりも死亡時満年齢の $\boxed{}$ が $\boxed{}$ 傾向にある．

最初の文の場合，良く言われる「女性は男性よりも長生きである」と同じ意味で $\boxed{\text{低い}}$ とすることができるでしょう．実際，平均値のみならず箱全体の位置を見たと

[4] 横向きの箱ひげ図の場合，左端に ⊢，右端に ⊣ を付す，ということです．

[5] 「おさまらない」ようなデータの例を，ぜひ試行錯誤して作ってみて下さい．

[6] 生命表を良く調べてその理由を探ってみることは意味のあることだと思います．平均値としての平均寿命を下げるのに寄与していると考えられるのはどの部分でしょうか．例えば，$d_0 = 184$ に対応する男 184 人が仮に生命表に従わずそれぞれ 100 歳まで生きたとしたら，平均寿命は 0.184 年 (= 100 年 × 184/10 万) 上昇することになりますが，中央値は変わりません．

きに，男の箱は女の箱よりも左にあります．もう一つはデータの散らばり（ばらつき）に関する文と考えてみて下さい．そうするとどうでしょうか．見た目だけでなく定量的に考察しましょう．2 つの箱の幅はどうなっているでしょうか．言い換えると，差 (第 3 四分位数) − (第 1 四分位数) の値です．それは**四分位範囲**と呼ばれる値です．表から直ちに，男の場合の値は $15(=90-75)$ で，女の場合の値は $12(=95-83)$ ですので，男のデータの方が $\boxed{\text{散らばり}}$ が $\boxed{\text{大きい}}$ 傾向にあると言えます．[7]

３．生命表を使って　— 確率（特に「条件付き確率」）の応用として —

生命表は死亡時満年齢という確率変数の**確率分布**を与えている，といった見方ができます．むしろそれが本来の目的を果たすために必要な見方であると言えます．以下では，それからどういった洞察が可能かを解説します．

(3-1) 生存率・死亡率とは

既に述べた生存数・死亡数の欄の右に「生存率」「死亡率」の欄があります．記号ではそれぞれ p_x，q_x と記されるものです．[8] 年齢を特定して述べた方がわかりやすいし，それで十分でしょうから，例えば 50 歳（男）の場合，つまり男の表で $x=50$ の行を考えます．50 歳の生存率とは 50 歳の人が 1 年後に生存している確率であり，それは 50 歳の生存数に占める 51 歳の生存数の割合として求められます：

$$(50\text{ 歳（男）の生存率}) = p_{50} = \frac{l_{51}}{l_{50}} = \frac{96712}{96948} = 0.99757$$

また，50 歳の死亡率とは 50 歳の人が 1 年後に死亡している確率であり，それは 50 歳の生存数に占める 50 歳の死亡数の割合として求められます：

$$(50\text{ 歳（男）の死亡率}) = q_{50} = \frac{d_{50}}{l_{50}} = \frac{236}{96948} = 0.00243$$

これら 2 つの数値の和が 1 となることは,「1 年後に生存している」という事象と「1 年後に死亡している」という事象とが互いに**余事象**の関係にあることの反映です．さらに強調しておきたいのは，分母が 10 万ではなく，これらの確率は実際には**条件付き確率**であることです．[9] それは，生存率については「50 歳で生存しているとき，その 1 年後も生存している確率」，死亡率については「50 歳で生存しているとき，その 1 年後に死亡している確率」と言い換えると理解できるのではないでしょうか．[10]

このように「x 歳の人が ⋯ 」という形でその主体の年齢を特定することは「x 歳で生存している」という条件の下での条件付き確率を考えることを意味します．そ

---

[7] こちらの傾向についてはあまり言われることがないように思いますが，やはり経験的にある程度感じ取れることかも知れません．なお，散らばりの度合いの別の指標である**分散**の値を求めるのに，生命表のエクセルファイルが利用できるはずです．もし分散の値を求めたら，**標準偏差**（分散の正の平方根）と**四分位偏差**（四分位範囲の半分の値）を比較してみて下さい．

[8] ある授業で出席者に生命表を配布して「自分の生存率に最も近い生存率を持つ異性は何歳か調べよ.」という課題を与えたことがあります．これを調べるだけでも，性別による違いがかなり印象付けられるのではないかと思います．

[9] 条件付きでない確率の例を述べておくと,（男が）88 歳で死ぬ確率であり，それは $d_{88}/100000=0.04133$ と求まります．

[10] 大学以降の「確率論」関係の講義で学ぶ際には，さらに明確に「50 歳で生存しているという条件の下での，その 1 年後も生存している条件付き確率」といった述べ方がなされることがあります．

れはまた，いわゆる全体集合を元々の 10 万人の集団から x 歳の生存者達 l_x 人に制限された集団に切り替える，といった操作でもあります．

(3-2) 50 年後の生存率・死亡率を求めよう．

授業のタイトルに掲げていた確率を求めないわけにはいきません．[11] 先ほど求めたのは 1 年後の生死に関する確率だけでしたが，より一般的な確率を求めるための情報が生命表には備わっていることに気づいた人も多いと思います．「とりあえず自分の場合にやってみて」という声が聞こえてきそうですが，私の場合に計算してみたところ，50 年後に生きている確率は非常に小さい値だとわかり，寂しくなりました．そこで，かなり若作りをさせてもらって，40 歳（男）の場合に求めてみます：

$$(40\text{ 歳（男）の 50 年後の生存率}) = \frac{l_{40+50}}{l_{40}} = \frac{l_{90}}{l_{40}} = \frac{28082}{98388} = 0.28542$$

$$(40\text{ 歳（男）の 50 年以内死亡率}) = 1 - 0.28542 = 0.71458$$

皆さんは自分の場合に求めてみて下さい．また，帰宅後，家族の方についても類似の確率を計算してあげてもよいと思います．

(3-3) 簡単な死亡保険を考えて，保険料を算出してみよう．

このように様々な確率が計算できるわけですが，最後にその応用について述べたいと思います．それは生命保険の保険料算出です．歴史的にも，生命表の存在理由の一つと言ってよいものです．最も簡単な生命保険として，1 年契約の死亡保険を考えます．やはり具体的な問題設定をしてみます：

・38 歳の女性が 1 年以内に死亡した場合，保険<u>金</u>100 万円が支払われる．
・この契約の対価としての保険<u>料</u>はいくらか？

明らかに 100 万円が実際に支払われる可能性を考慮する必要があります．それは 38 歳（女）の死亡率 $q_{38} = 0.00050$ に等しいと考えることで，保険料は 100 万円 $\times 0.00050 =$ 500 円 と算出されます．一言で言えば**期待値**の計算です．保険会社が支払うことになる金額にその確率が乗じられています．(形式上は，この 38 歳の女性が 1 年後も生存している場合には 0 円を支払うものとして考え，それに生存率 $p_{38} = 0.99950$ が乗じられた結果としての 0 円も加算されていると考えなければなりません．)

　このような簡潔な解答だけでは天下り過ぎてピンと来ないと思いますので，これからその妥当性を契約者（正確には被保険者）集団のモデルを作って見てみましょう．例えば，上記の死亡保険の契約者が同時に 1 万人いたと仮定します．保険会社の立場に立ってみましょう．そうすると，一人につき保険料 500 円を徴収することで

$$\boxed{\text{保険料収入総額}} = 500\text{ 円} \times 1\text{ 万} = 500\text{ 万円}$$

となります．一方，1 年後の保険金支出の状況はどうなると考えられるでしょうか．こちらには，契約者 1 万人中で死亡が何件発生するか，という肝心の点に偶然性が

[11] 公正を期すために言及しておくと，「あなた」についての確率ではなく，「この 10 万人の集団から無作為に選ばれた一人」についての確率です．

作用し，それから免れるわけにはいきません．それでも，死亡率によりその見込みを与えることはできます．この問題の場合だと $q_{38} = 0.00050$ より

$$(1\text{年以内死亡数（見込み）}) = 1\text{万人} \times 0.00050 = 5\text{人}$$

であり，この 5 人にそれぞれ保険金 100 万円を支払うと考えることで次を得ます．

$$\boxed{\text{保険金支出総額（見込み）}} = 100\text{万円} \times 5 = 500\text{万円}$$

こうして示された収入と支出（見込み）との釣り合いは，500 円という保険料設定の合理性に根拠を与えてくれます.（現実には「見込み」が悪い方にはずれる可能性もありますので，そのようなリスクへの備えも必要となります.）皆さんも，性別・年齢・保険金を各自で設定して，1 年契約の死亡保険の保険料を算出してみて下さい．

これまでに述べたことから，契約期間をもっと長くする等して様々な生命保険を考えることが可能であることは，もはや想像に難くないでしょう．ただそのような場合，徴収した保険料総額の長期的運用についても考慮する必要性が生じます．それは現実的かつ重要な要素ですが，基本的に複利計算により扱うことができ，高校で学ぶ等比数列の考え方が本質的な役割を果たします．

このように，数学を利用して現実的な問題に対処しようとする際，必要となる数学の範囲を予め限定するのは困難ですし，限定した場合には対処できる問題も必然的に限られてしまうことになります．その意味でも，在学中は高校数学を幅広くしっかりと修得しておいて欲しいと思います．

４．その他

世の中で起きているでき事を，論理的思考を通じて考察してくれたらと思います．さらに，可能な場合にはデータや数学的手法を適切に利用してくれるよう望みます．

私見ですが，5 年後に発表される予定の第 24 回生命表は，今回利用した第 23 回生命表と比べて様相が異なってくるのではないかと予想しています．実際，毎年作成されている「簡易生命表」の最新版[12] では，平均寿命が男は 0.09 年，女は 0.14 年，前年版を下回りました．[13] また，この報道に続いて，ほとんどの都道府県で人口が減少していることが総務省の人口動態調査で示されているとの報道もありました．

個人的に非常にわかりやすいと感じている関連資料を紹介しておきます．それは，総務省統計局[14] の「人口推計 2022 年（令和 4 年）7 月報」p.4 に掲載されている「総人口の推移」のグラフです．これから窺うことができる状況を記しておきます．

- 2020 年の 2 月前後から総人口の増減の振幅が小さくなった．
- 2021 年の 2 月前後から総人口の減少の速さが大きくなった．
- 2022 年の 2 月前後から総人口の減少の速さがさらに大きくなった．

これらはそれぞれ何に起因するのか．この問に真摯に向き合うことを宿題とします．

[12] 令和 3 年簡易生命表　https://www.mhlw.go.jp/toukei/saikin/hw/life/life21/index.html

[13] 男の場合だと 10 万人の生存年数合計で 9 千年の減少に相当します．例えば，(脚注 6 と同様の思考実験に過ぎませんが) 10 万人中 150 人の死亡時満年齢が前年版簡易生命表と比べて一律に 60 年低ければ，そうした平均寿命の低下が生じます．

[14] 「人口推計」のページ　https:www.stat.go.jp/data/jinsui/index.html