

割合の近似値がいくつもある場合に 共通分母を探索するコマンド

アンケート調査や疫学報告において簡便な
記載ミス発見とサンプルサイズ推定のために

2022年10月12日(金) 下野寿之

本文書の構成は、かなり学術よりの人が読むことを意識している。IT勉強会などで話すためのスライド資料としては、さらに平易な表現と大幅な文書全体の再構成を要するであろう。



1. 問題提起

割合を表す数の近似値が、いくつか与えられた場合に、分母を推定したいことがある。
まず、その具体例のいくつかと問題意識を提起する。

素朴な疑問(1/2)

<https://www.mhlw.go.jp/content/000940528.pdf>
東京都のワクチン接種会場で配布資料の一部

■ 接種後7日間に現れた症状

報告割合	1回目接種後	2回目接種後	3回目接種後
50%以上		圧痛 65.2%	圧痛 81.4% 疲労 63.3% 疼痛 54.6% 筋肉痛 51.0%
10 - 50%	圧痛 48.2% 疼痛 26.9% 疲労 23.1% 頭痛 21.6% 筋肉痛 20.0% 倦怠感 12.2%	疼痛 45.6% 疲労 35.6% 筋肉痛 30.8% 頭痛 29.6% 倦怠感 26.4% 関節痛 14.8%	倦怠感 46.9% 頭痛 45.9% 関節痛 28.6% 発熱 17.3% 恶心・嘔吐 13.3% 腫脹・硬結 11.3% 紅斑 10.3%
1 - 10%	関節痛 6.7% 恶心・嘔吐 5.9% 発熱 2.4%	恶心・嘔吐 7.2% 腫脹・硬結 5.6% 紅斑 4.8% 発熱 4.4%	

(注) 対象:1回目接種後253人、2回目接種後250人、3回目接種後97人

出典:審査報告書より改編

数値が10個程度以上あれば、
確認作業は容易ではなく、
よく記載ミスは発生する。

上記の割合は正しく記載されているか?
間違いがあるとしたら、 ありうる理由は?

素朴な疑問(2/2)

<https://nlab.itmedia.co.jp/nl/articles/1812/12/news047.html>
バンダイのお年玉についての記事(2018年)

小学生 TOP5			中学生 TOP5		
1位	貯金	39.9%	1位	貯金	33.7%
2位	ゲーム機・ゲームソフト	30.4%	2位	文房具、雑貨	31.8%
3位	おもちゃ、カードゲーム(ゲーム機・ゲームソフト除く)	29.5%	3位	衣服・衣類雑貨	23.5%
4位	書籍(マンガ以外)	18.0%	4位	ゲーム機・ゲームソフト	22.7%
5位	お菓子やジュースなどの飲食物	17.2%	5位	マンガ	22.0%

男子 TOP5 (小中学生)			女子 TOP5 (小中学生)		
1位	ゲーム機・ゲームソフト	37.5%	1位	貯金	39.4%
2位	貯金	36.2%	2位	文房具、雑貨	35.7%
3位	おもちゃ、カードゲーム(ゲーム機・ゲームソフト除く)	27.3%	3位	書籍(マンガ以外)	21.5%
4位	お菓子やジュースなどの飲食物	18.8%	4位	衣服・衣類雑貨	20.7%
5位	マンガ	18.5%	5位	お菓子やジュースなどの飲食物	18.4%

複数回答 有効回答726人

どれかの内訳の人数が少なすぎて、その場合に信頼性の低い統計値になっていないだろうか?

「有効回答人数は726人」というヒントあり。
小学生、中学生、男子、女子のそれぞれの人数は?

別の疑問

複数の無理数を 共通する分母の分数で
誤差をたとえば0.5%以内で近似したい。

- 例: 2,3,5,7,8の各平方根に対してどうだろうか?
→ 最小解は60となる。

```
> denomfind -y, -g10 -D-6 -a0.005 `perl -E'say sqrt($_) for 2,3,5..8'` | expandtab
6 ratios are given. 10 denominators have found up to 98. (denomfind)
denom fit f1:1.4142135623731 f2:1.73205080756888 f3:2.23606797749979 f4:2.44948974278318 f5:2.64575131106459 f6:2.82842712474619
60   6 85/60-0.002453 104/60-0.001283 134/60+0.002735 147/60-0.000510 159/60-0.004249 170/60-0.004906
82   6 116/82-0.000421 142/82+0.000343 183/82+0.004361 201/82-0.001730 217/82-0.000590 232/82-0.000841
83   6 117/83+0.004575 144/83-0.002889 186/83-0.004896 203/83+0.003707 220/83-0.004851 235/83-0.002898
85   6 120/85+0.002449 147/85+0.002639 190/85+0.000774 208/85+0.002431 225/85-0.001308 240/85+0.004898
92   6 130/92+0.001170 159/92+0.003790 206/92-0.003062 225/92+0.003838 243/92+0.004447 260/92+0.002340
94   6 133/94-0.000680 163/94-0.001992 210/94+0.002025 230/94+0.002681 249/94-0.003185 266/94-0.001360
95   6 134/95+0.003687 165/95-0.004791 212/95+0.004489 233/95-0.003142 251/95+0.003646 269/95-0.003152
96   6 136/96-0.002453 166/96+0.002884 215/96-0.003515 235/96+0.001573 254/96-0.000082 272/96-0.004906
97   6 137/97+0.001842 168/97+0.000092 217/97-0.001045 238/97-0.004119 257/97-0.003733 274/97+0.003685
98   6 139/98-0.004154 170/98-0.002643 219/98+0.001374 240/98+0.000510 259/98+0.002894 277/98+0.001897
> []
```

引き続く疑問

1. 複数の割合近似値から共通分母の値を求めたい。
2. 共通分母の候補値は、どんな現れ方をするか？
3. そこにある未知の数理的な性質はないか？
 - 何桁の近似値が何個揃えば、分母を推定しやすいか？
4. あらうる手入力ミス等への良い対応の仕方は？
5. 上記の解決策に対して、有用性の確認方法は？





2. プログラムの設計

割合の数を表す近似が複数ある場合に、
その分母の候補値を算出するプログラムを作ったので、
そのプログラムの設計について解説する。

計算原理

1. 近似値Q を、実数上の半区間と見なす。
 - Qは小数点以下 d 桁まで10進数で記載されていたとする。
 - その半区間の幅は 0.1^d (0.1のd乗)となる。
2. Qの近似前の数は、丸め方で異なる区間となる。
 - $[Q, Q + 0.1^d)$: **切り捨て** — 左は閉端、右は開端の半区間
 - $(Q - 0.1^d, Q]$: **切り上げ** — 左は開端、右は閉端の半区間
 - $[Q - 0.5 \times 0.1^d, Q + 0.5 \times 0.1^d)$ (#) : (単純な1回だけの)**四捨五入**
 - $[Q - 0.55 \times 0.1^d, Q + 0.5 \times 0.1^d)$: **2回目の四捨五入で1桁縮めた場合**
 - **偶数丸め**の場合、Qの最下位の偶/奇により区間(#)を閉/開に変える。
3. 候補となる分母 $D=1,2,3,\dots \in \mathbb{N}$ に対し:
 - その区間にDを乗算。その区間に含まれる整数N全てが**分子の候補**。
 - 10進数を意識した計算を要する(2進小数で計算すると誤差が発生)。
4. 与えた全ての割合の近似値に対して:
 - $D=1,2,3,\dots \in \mathbb{N}$ のそれぞれについて、分子の候補が存在するか?
 - 前項目を目で見て分かりやすく、途中の計算結果も含めて表示する。

参考用；知っている各種ソフトを比較する意図が無いと、このページはあまり役に立たないであろう。

設計方針

- 汎用性 ← 古めの環境でも、50年後でも。
- コマンド入力時に使えるオプション
 - 的確なオプションを設計。結果的に多数。
- 出力結果:
 - 結果視認性の確保のため、色(ANSIカラー)を多用。
- コマンドラインで使うことを前提とする
 - なお、JavaScript + HTMLも検討したが、未着手。
- 機能の複雑さに応じた内部設計
 - 必ずしも、平易なプログラムでは無い。
- 切れ味の良さ ← 特に、素朴な疑問に的確に応じているか

コマンドのインストール方法

- インストール: **cpanm App::denomfind**
- アンインストール: **cpanm -U App::denomfind**

cpanmが使えない場合は、**cpan**でも良いが、モジュールのアンインストールが難しくなる。

上記は Perlのモジュールのインストールでよく用いられる方法である。

cpanmやcpanによる一般的なインストールの方法は、下記で網羅的に説明をしている。

1. <https://metacpan.org/release/TULAMILI/Table-Hack-0.12/source/Table-Hack.pdf>
2. <https://www.ite.or.jp/wp/wp-content/uploads/opendata/tool201807.pdf>

上記2個のURLにある同じPDF形式ファイルは、2018年に下野寿之が用意している。

コマンドの用途と有用性：

- 分母となる整数の候補を見つける。
 - 割合の近似値から、求まる。
 - 割合近似値が多数あれば、同じ分母なら1個に定まり得る。
- 割合数値を記載した文書の**完成度**が分かる:
 1. 数値の計算の仕方について、文書の解説に間違いはないか?
 2. 近似の計算に不備は無いか? (同じ数を2回近似するなど)
 3. 割合を計算するときに、分母が一部だけ異なることは無いか?
 4. Nまたは割合数値の記載ミスは、十分少ないか?
 5. サンプルサイズNが過小でないか?
 6. そもそもたらめな数を記載していないか?
- 上記で発見した不備は、**訂正/改善**が可能な場合も多い。
 - 難解な文書の読解を容易にしたり、誤解を正すこともできる場合もある。

参考用；コマンドを何度も使わない無いと、このページはあまり役に立たないであろう。

denomfindを使ったコマンド文の文法

- **denomfind** (options) arg_1 arg_2 .. arg_n
 - g [最小分母,]探索個数
 - y 「整数分子を持ちうる割合の個数」のフィルタリングの指定
 - D num ← シアン色で、割り算の式と小数点以下num桁を表示
 - I ← 分子になりうる値の区間範囲を出力する指定
 - Q ← 分子の整数の候補をコンマ区切りに。(簡便な出力となる。)
 - % ← 指定する割合の近似値が、百分率(%)の数値であることの指定
 - arg_1 arg_2 .. arg_n ← 割合の近似値 (必須の入力値)
- その他のオプション：
 - -f は切り捨て、 -c は切り上げによる近似であることを指定
 - -f か -c の指定が無い場合は、 **四捨五入**による近似を指定
 - -52 により、 四捨五入した値の最下位の桁を四捨五入したことを指定
 - -5e は偶数丸め、 -5o は奇数丸めであることを指定

※ 出力に関するオプションは英語大文字を採用した。

コマンド denomfind の出力形式

出力各行は、多数の列でタブ区切りで構成される。

- 1列目 : **候補となる分母** の値 $D \in \mathbb{N}$
- 2列目 : 与えた近似値の内、次を満たす個数:
「 D を分母として、分子を整数にすることが可能」
- 3列目以降 (オプションにより異なる出力を選択できる):
 - 分数の候補を表す、**割り算の式**。 $(-D\ num)$
 - 分子候補値の、**整数に限らない実数の区間**の表示。 $(-I)$
 - 分子になりうる整数全てをコンマ区切りで。 $(-Q)$

出力は、ANSIカラー付きのテキスト形式である。

- 別コマンドを使い **colorplus -0** で、色無し平文テキストへ変換。

denomfind と 共に使うプログラム

- ある分母の値が候補なら、
その値の倍数も全て候補になるので、
それらの出力を最初だけに抑制する
コマンド **primefind** も作成した。
- App::expandtab に含まれる **expandtab** も便利。
タブ文字区切り入力をタブで立て揃え出力する。
- 出力の色を消す**colorplus -0**(App::colorplus)
- テキストビューアの **less** を
使いこなしていることは重要である。

コマンドdenomfindの完成度について

- 4年前作成のプログラムについて、重大なバグを最近発見(1.0%を与えても1%とみなして近似幅が拡大)。
- 200行弱のPerlプログラムをほぼ全部リファクタをして点検。
- さらに新たな機能をオプションの形で追加。出力表の形式も何度も形を変えて、見やすくした。
- 手を新たに加えた部分は、再度ほぼ全部リファクタして点検する必要はありそう。新しい機能を加えると、それ以外の関係する古い機能に対して、バグを見つけることがある。

コマンドdenomfindの今後の計画

- 他の言語に移植したい。
 - 数理的な思いつきを検証容易としたい。
 - 必要な機能の網羅と洗練を済ませることが前提。
- ゆくゆくは、普及させたい。
 - できるだけ多くの人に使ってもらいたい。
 - ジャーナリストの人も容易に使えるようにしたい。
 - 日本語の部分を英語化したい。
- そのために、機能の設計は高度に洗練したい。



3. コマンドの実行例

実際の実行画面で、どのような操作画面になるか、
実感に近いイメージを、一旦、提供する。

コマンドの使用例

異なる5個の条件に「**2.0% 50.7% 13.5% 21.6% 6.8%**」が該当した時に、サンプルサイズを**148人**が正しいかどうか確かめたい。

denom	fit	f1:0.020	f2:0.507	f3:0.135	f4:0.216	f5:0.068
141	1	-	-	19/141=.13475	-	-
142	1	-	72/142=.50704	-	-	-
143	0	-	-	-	-	-
144	1	-	73/144=.50694	-	-	-
145	0	-	-	-	-	-
146	2	-	74/146=.50685	-	-	10/146=.06849
147	2	3/147=.02041	-	-	-	10/147=.06803
148	5	3/148=.02027	75/148=.50676	20/148=.13514	32/148=.21622	10/148=.06757
149	1	3/149=.02013	-	-	-	-
150	2	3/150=.02000	76/150=.50667	-	-	-
151	1	3/151=.01987	-	-	-	-
152	2	3/152=.01974	77/152=.50658	-	-	-
153	2	3/153=.01961	-	-	33/153=.21569	-
154	0	-	-	-	-	-
155	1	-	-	21/155=.13548	-	-
156	1	-	-	21/156=.13462	-	-
157	0	-	-	-	-	-
158	0	-	-	-	-	-
159	0	-	-	-	-	-
160	0	-	-	-	-	-

分母の値を141から初めて、20個の場合について結果を表示した。

コマンドの使用例

5個の条件に「**2.0% 50.7% 13.5% 21.6% 6.8%**」が該当した場合に、サンプルサイズを推定(5個中2個以内の記載ミスの場合も考えた)。

```
> denomfind -y 3..5 -g141,20 -D4 -% 2.0 50.7 13.5 21.6 6.8 | expandtab
5 ratios are given. 20 denominators have found up to 349. (denomfind)
denom fit f1:0.020 f2:0.507 f3:0.135 f4:0.216 f5:0.068
148 5 3/148=.0203 75/148=.5068 20/148=.1351 32/148=.2162 10/148=.0676
205 3 4/205=.0195 104/205=.5073 - - 14/205=.0683
207 3 - 105/207=.5072 28/207=.1353 - 14/207=.0676
222. 3 - - 30/222=.1351 48/222=.2162 15/222=.0676
245 3 5/245=.0204 - 33/245=.1347 53/245=.2163 -
250 3 5/250=.0200 - - 54/250=.2160 17/250=.0680
251 3 5/251=.0199 - 34/251=.1355 - 17/251=.0677
278 3 - 141/278=.5072 - 60/278=.2158 19/278=.0683
282 3 - 143/282=.5071 38/282=.1348 61/282=.2163 -
292 3 - 148/292=.5068 - 63/292=.2158 20/292=.0685
294 3 6/294=.0204 149/294=.5068 - - 20/294=.0680
296. 5 6/296=.0203 150/296=.5068 40/296=.1351 64/296=.2162 20/296=.0676
304 3 6/304=.0197 154/304=.5066 41/304=.1349 - -
306 3 6/306=.0196 155/306=.5065 - 66/306=.2157 -
310 3 - - 42/310=.1355 67/310=.2161 21/310=.0677
342 3 7/342=.0205 - 46/342=.1345 74/342=.2164 -
343 3 7/343=.0204 174/343=.5073 - 74/343=.2157 -
347 4 7/347=.0202 176/347=.5072 47/347=.1354 75/347=.2161 -
348 3 7/348=.0201 - 47/348=.1351 75/348=.2155 -
349 3 7/349=.0201 177/349=.5072 47/349=.1347 - -
> █
```

分母の少ない順に「5個の割合が整数比を表す場合」を3～5個の場合に限定し、20個探索。

コマンドの使用例

表示内容は、割り算式ではなくて、分子の値の範囲を表示。
緑色の場合は、その区間に整数が1個以上含まれることを示す。

```
> denomfind -y0,3,4,5 -g290,20 -I -% 2.0 50.7 13.5 21.6 6.8 | expandtab
5 ratios are given. 20 denominators have found up to 351. (denomfind)
denom fit numerators_1      numerators_2      numerators_3      numerators_4      numerators_5
292   3 [5.694 5.986) [147.898 148.19) [39.274 39.566) [62.926 63.218) [19.71 20.002)
294   3 [5.733 6.027) [148.911 149.205) [39.543 39.837) [63.357 63.651) [19.845 20.139)
296.  5 [5.772 6.068) [149.924 150.22) [39.812 40.108) [63.788 64.084) [19.98 20.276)
304   3 [5.928 6.232) [153.976 154.28) [40.888 41.192) [65.512 65.816) [20.52 20.824)
306   3 [5.967 6.273) [154.989 155.295) [41.157 41.463) [65.943 66.249) [20.655 20.961)
310   3 [6.045 6.355) [157.015 157.325) [41.695 42.005) [66.805 67.115) [20.925 21.235)
313   0 [6.1035 6.4165) [158.5345 158.8475) [42.0985 42.4115) [67.4515 67.7645) [21.1275 21.4405)
314   0 [6.123 6.437) [159.041 159.355) [42.233 42.547) [67.667 67.981) [21.195 21.509)
316   0 [6.162 6.478) [160.054 160.37) [42.502 42.818) [68.098 68.414) [21.33 21.646)
317   0 [6.1815 6.4985) [160.5605 160.8775) [42.6365 42.9535) [68.3135 68.6305) [21.3975 21.7145)
321   0 [6.2595 6.5805) [162.5865 162.9075) [43.1745 43.4955) [69.1755 69.4965) [21.6675 21.9885)
330   0 [6.435 6.765) [167.145 167.475) [44.385 44.715) [71.115 71.445) [22.275 22.605)
331   0 [6.4545 6.7855) [167.6515 167.9825) [44.5195 44.8505) [71.3305 71.6615) [22.3425 22.6735)
332   0 [6.474 6.806) [168.158 168.49) [44.654 44.986) [71.546 71.878) [22.41 22.742)
342   3 [6.669 7.011) [173.223 173.565) [45.999 46.341) [73.701 74.043) [23.085 23.427)
343   3 [6.6885 7.0315) [173.7295 174.0725) [46.1335 46.4765) [73.9165 74.2595) [23.1525 23.4955)
347   4 [6.7665 7.1135) [175.7555 176.1025) [46.6715 47.0185) [74.7785 75.1255) [23.4225 23.7695)
348   3 [6.786 7.134) [176.262 176.61) [46.806 47.154) [74.994 75.342) [23.49 23.838)
349   3 [6.8055 7.1545) [176.7685 177.1175) [46.9405 47.2895) [75.2095 75.5585) [23.5575 23.9065)
351   3 [6.8445 7.1955) [177.7815 178.1325) [47.2095 47.5605) [75.6405 75.9915) [23.6925 24.0435)
> |
```

コマンドの使用例

2.0% 13.5% 21.6% の3個の近似値が与えられた場合に、分母が150以下でこれらの割合を実現するものを探した例。

```
> denomfind -y1..3 -D5 -I -Q -g150,-inf -% 2.0 13.5 21.6 | expandtab
3 ratios are given. 28 denominators have found down to 1. (denomfind)
denom fit numerators_1 numerators_2 numerators_3 dividends
150. 1 [2.925 3.075) [20.175 20.325) [32.325 32.475) 3 prime
149 1 [2.9055 3.0545) [20.0405 20.1895) [32.1095 32.2585) 3 prime
148 3 [2.886 3.034) [19.906 20.054) [32/148=.21622 [31.894 32.042) 3,20,32
147. 1 [2.8665 3.0135) [19.7715 19.9185) [31.6785 31.8255) 3 prime
141 1 [2.7495 2.8905) [19/141=.13475 [18.9645 19.1055) [30.3855 30.5265) 19 prime
139 1 [2.7105 2.8495) [18.6955 18.8345) [29.9545 30.0935) 30 = 2 x 3 x 5
134 1 [2.613 2.747) [18.023 18.157) [28.877 29.011) 29 prime
133 1 [2.5935 2.7265) [18/133=.13534 [17.8885 18.0215) [28.6615 28.7945) 18 = 2 x 3^2
126 1 [2.457 2.583) [17/126=.13492 [16.947 17.073) [27.153 27.279) 17 prime
125 1 [2.4375 2.5625) [16.8125 16.9375) [27/125=.21600 [26.9375 27.0625) 27 = 3^3
116 1 [2.262 2.378) [15.602 15.718) [25/116=.21552 [24.998 25.114) 25 = 5^2
111. 2 [2.1645 2.2755) [15/111=.13514 [14.9295 15.0405) [24/111=.21622 [23.9205 24.0315) 15,24 (9q+6)
104. 1 [2.028 2.132) [14/104=.13462 [13.988 14.092) [22.412 22.516) 14 = 2 x 7
102. 2 [1.989 2.091) [13.719 13.821) [22/102=.21569 [21.981 22.083) 2,22 (20q+2)
101 1 [1.9695 2.0705) [13.5845 13.6855) [21.7655 21.8665) 2 prime
100. 1 [1.95 2.05) [13.45 13.55) [21.55 21.65) 2 prime
99 1 [1.9305 2.0295) [13.3155 13.4145) [21.3345 21.4335) 2 prime
98. 1 [1.911 2.009) [13.181 13.279) [21.119 21.217) 2 prime
97 1 [1.8915 1.9885) [13.0465 13.1435) [20.9035 21.0005) 21 = 3 x 7
96 1 [1.872 1.968) [12.912 13.008) [20.688 20.784) 13 prime
89 1 [1.7355 1.8245) [11.9705 12.0595) [19.1795 19.2685) 12 = 2^2 x 3
88 1 [1.716 1.804) [11.836 11.924) [18.964 19.052) 19 prime
74. 2 [1.443 1.517) [10/74=.13514 [9.953 10.027) [15.947 16.021) 10,16 (6q+4)
52 1 [1.014 1.066) [6.994 7.046) [11.206 11.258) 7 prime
51 2 [0.9945 1.0455) [6.8595 6.9105) [10.9905 11.0415) 1,11 (10q+1)
50 1 [0.975 1.025) [6.725 6.775) [10.775 10.825) 1
49 1 [0.9555 1.0045) [6.5905 6.6395) [10.5595 10.6085) 1
37 2 [0.7215 0.7585) [4.9765 5.0135) [7.9735 8.0105) 5,8 (3q+2)
>
```



4. 冒頭2個の題材への適用

2回目接種の250人という数
については、正しそうだ。

■ 接種後7日間に現れた症状

報告割合	1回目接種後	2回目接種後	3回目接種後
50%以上		圧痛 65.2%	圧痛 81.4% 疲労 63.3% 疼痛 54.6% 筋肉痛 51.0%
10 - 50%	圧痛 48.2% 疼痛 26.9% 疲労 23.1% 頭痛 21.6% 筋肉痛 20.0% 倦怠感 12.2%	疼痛 45.6% 疲労 35.6% 筋肉痛 30.8% 頭痛 29.6% 倦怠感 26.4% 関節痛 14.8%	倦怠感 46.9% 頭痛 45.9% 関節痛 28.6% 発熱 17.3% 恶心・嘔吐 13.3% 腫脹・硬結 11.3% 紅斑 10.3%
1 - 10%	関節痛 6.7% 恶心・嘔吐 5.9% 発熱 2.4%	恶心・嘔吐 7.2% 腫脹・硬結 5.6% 紅斑 4.8% 発熱 4.4%	

(注) 対象:1回目接種後253人、2回目接種後250人、3回目接種後97人

出典:審査報告書より改編

ここに書かれた説明は、1回目接種後と3回目接種後のことについて、誤りであると分かる。

1回目が253人としても、～%表示の9個の内、5個しか分子が整数になり得ない。255の書き間違いの可能性もあるが、それでも、8個だけが整数分子。

```
> denomfind -y5.. -D0 -g10 -% 48.2 26.9 23.1 21.6 20.0 12.2 6.7 5.9 2.4 | expandtab
9 ratios are given. 10 denominators have found up to 328. (denomfind)
denom fit f1:0.482 f2:0.269 f3:0.231 f4:0.216 f5:0.200 f6:0.122 f7:0.067 f8:0.059 f9:0.024
208 5 - 56/208 48/208 45/208 - - 14/208 - 5/208
238 5 - 64/238 55/238 - - 29/238 16/238 14/238 -
245 6 118/245 66/245 - 53/245 49/245 30/245 - - 6/245
253 5 122/253 68/253 - - - 17/253 15/253 6/253
255 8 123/255 - 59/255 55/255 51/255 31/255 17/255 15/255 6/255
286 5 - 77/286 66/286 - - 35/286 - 17/286 7/286
290 5 - 78/290 67/290 - 58/290 - - 17/290 7/290
305 5 147/305 82/305 - 66/305 61/305 - - 18/305 -
320 6 - 86/320 74/320 69/320 64/320 39/320 - 19/320 -
328 5 158/328 - - 71/328 - 40/328 22/328 - 8/328
```

```
> denomfind -y4.. -D0 -g1,10 -% 81.4 63.3 54.6 51.0 46.9 45.9 28.6 17.3 13.3 11.3 10.3 | expandtab
11 ratios are given. 10 denominators have found up to 194. (denomfind)
denom fit f1:0.814 f2:0.633 f3:0.546 f4:0.510 f5:0.469 f6:0.459 f7:0.286 f8:0.173 f9:0.133 f10:0.113 f11:0.103
49 4 - 31/49 - 25/49 23/49 - 14/49 - - - -
97 4 79/97 - 53/97 - - - - - 11/97 10/97
98 7 - 62/98 - 50/98 46/98 45/98 28/98 17/98 13/98 - -
133 4 - - - - 61/133 38/133 23/133 - 15/133 -
145 4 118/145 - - 74/145 68/145 - - - - 15/145
147. 4 - 93/147 - 75/147 69/147 - 42/147 - - - -
150 4 - 95/150 - - - - 26/150 20/150 17/150 -
177 4 144/177 112/177 - - 83/177 - - - 20/177 -
185 5 - - 101/185 - - 85/185 53/185 32/185 - - 19/185
194 7 158/194 - 106/194 99/194 91/194 89/194 - - - 22/194 20/194
```

3回目が97人としても、～%の表示の内、4個しか分子が整数になり得ない。98人の場合の7個の項目の調査と合併した11個の項目かも知れない。

3回目接種後97人

3回目接種後

圧痛 81.4% 疲労 63.3%
疼痛 54.6% 筋肉痛 51.0%

倦怠感 46.9% 頭痛 45.9%
関節痛 28.6% 発熱 17.3%
恶心・嘔吐 13.3%
腫脹・硬結 11.3% 紅斑 10.3%

denomfind で計算した結果、
97人のデータと98人のデータの
混在の可能性が見つかった。

分母を75から102まで1個ずつ全て調べて、与えられた11個の割合数値を四捨五入(同じ数に四捨五入を2回行ったことも仮定した)で実現した場合にありうる整数比の分数をシアン色で表示している。2列目の薄い文字(1~7の数)で、何個が整数比の分数を実現したかを表す。

denom	fit	f1:0.814	f2:0.633	f3:0.546	f4:0.510	f5:0.469	f6:0.459	f7:0.286	f8:0.173	f9:0.133	f10:0.113	f11:0.103	dividends
75	2	-	-	-	-	-	-	-	13/75	10/75	-	-	13,10 (3q+1)
77	2	-	-	42/77	-	-	-	22/77	-	-	-	-	42,22 (20q+2)
78.	1	-	-	-	-	-	-	-	-	-	-	8/78	8 = 2^3
79	1	-	50/79	-	-	-	-	-	-	-	-	-	50 = 2 x 5^2
80	1	-	-	-	-	-	-	-	-	9/80	-	-	9 = 3^2
81	2	-	-	-	-	38/81	-	-	14/81	-	-	-	38,14 (24q+14)
83	1	-	-	-	-	-	-	-	11/83	-	-	-	11 prime
84.	1	-	-	-	-	-	24/84	-	-	-	-	-	24 = 2^3 x 3
85	1	-	-	-	-	-	39/85	-	-	-	-	-	39 = 3 x 13
86.	1	70/86	-	-	-	-	-	-	-	-	-	-	70 = 2 x 5 x 7
87.	1	-	-	-	-	-	-	-	-	-	9/87	-	9 = 3^2
88.	1	-	48/88	-	-	-	-	-	-	-	-	-	48 = 2^4 x 3
90.	2	-	57/90	-	-	-	-	-	12/90	-	-	-	57,12 (45q+12)
91.	1	-	-	-	-	-	26/91	-	-	-	-	-	26 = 2 x 13
96	2	-	-	49/96	45/96	-	-	-	-	-	-	-	49,45 (4q+1)
97	4	79/97	-	53/97	-	-	-	-	-	11/97	10/97	-	79,53,11,10
98	7	-	62/98	-	50/98	46/98	45/98	28/98	17/98	13/98	-	-	62,50,46,45,28,17,13
99.	1	-	-	54/99	-	-	-	-	-	-	-	-	54 = 2 x 3^3
100	1	-	-	-	51/100	-	-	-	-	-	-	-	51 = 3 x 17
102	2	83/102	-	-	52/102	-	-	-	-	-	-	-	83,52 (31q+21)

“小学生”で
あり得る人数:
471,516,529..

小学生 TOP5					中学生 TOP5				
1位	▷ denomfind -y5 -Qg3 -%20 \	39.9%	1位	▷ denomfind -y5 -Qg3 -%20 \	33.7%				
2位	39.9 30.4 29.5 18.0 17.2	30.4%	2位	33.7 31.8 23.5 22.7 22.0	31.8%				
3位	denom fit dividends	29.5%	3位	denom fit dividends	23.5%				
4位	471 5 188,143,139,85,81	18.0%	4位	255 5 86,81,60,58,56	22.7%				
5位	516 5 206,157,152,93,89	17.2%	5位	264 5 89,84,62,60,58	22.0%				
	▷			▷					

“男子”で
あり得る人数:
373,384,520..

男子 TOP5 (小中学生)					女子 TOP5 (小中学生)				
1位	▷ denomfind -y5 -Qg3 -%20 \	37.5%	1位	▷ denomfind -y5 -Qg3 -%20 \	39.4%				
2位	37.5 36.2 27.3 18.8 18.5	36.2%	2位	39.4 35.7 21.5 20.7 18.4	35.7%				
3位	denom fit dividends	27.3%	3位	denom fit dividends	21.5%				
4位	373 5 140,135,102,70,69	18.8%	4位	353 5 139,126,76,73,65	20.7%				
5位	384 5 144,139,105,72,71	18.5%	5位	381 5 150,136,82,79,70	18.4%				
	▷			▷					

複数回答 有効回答726人

“中学生”で
あり得る人数:
255,264,409..

“女子”で
あり得る人数:
353,381,526..

- 有効回答数726人とされている小中学生のお年玉について、小学生/中学生/男子/女子の内訳でそれぞれの使い道のトップ5項目の割合が表示されていた。
- その割合の数を四捨五入と見なし、5個の数を整数比の分数である場合に、**各内訳の5個に共通する分母としてあり得る数**を、小さい順に算出した。
- **denomfind -y5 -Q -g3 -% -20 39.9 30.4 29.5 18.0 17.2** のようなコマンドを用いた。
- 39.9などの数を-%で%表記と見なし、-y5により5個の割合数値が全て5個とも整数分子が見つかる整数分母を探索した。
- なお、-Qで分子の値を表示して、-20により冗長な2次情報の抑制を指定している。
- それらのあり得る数の内、小学生と中学生で足して726、男子と女子で足しても726となる数は、上記の丸で囲った数のみである。
- **各内訳(小/中/男/女)のどれもが、255人以上の少なくない人数と判明。**
- 候補として算出した数4個は各最小値だけが解となつた不思議さがある。

“割合の数値の不具合”がある場合の類型

1. 単純に数値の書き写しをミスした：
 - 例：十数個もあったので、どこか間違う。
2. 数値丸めの際が不適切であった：
 - 四捨五入を2回(以上)行った：
例: $3.45 \rightarrow 3.5 \rightarrow 4$; 3.45 の単純な四捨五入は 3 。
このような場合は、 $1/20$ の確率で、割合近似値が間違う。
 - 切り捨てと四捨五入の混在。
3. データの作られ方の理解を文書作成者が誤った：
 - 例. 異なるサイズの集団の結果を併合している。
4. そもそもデータが出鱈目な場合：
 - 文書の体裁だけを取り繕う人が、文書作成に関わっていた。

5. 補足

謎な現象とその考察

- ある場合に7個もの分子の数が全て3の倍数であった。
- 珍しくて、特記すべき事象と言えるか？
- 全ての分子が、特殊な性質を満たす場合、
 p 値のような考え方で、その場合の分母は、特殊だと言えるだろうか？— 他、整数 q に対して $5q+3$ のような整数が7個もある場合

```
> denomfind -g300 -Q -y7.. -% 65.2 45.6 35.6 30.8 29.6 26.4 14.8 | \
grep "q" | expandtab
7 ratios are given. 300 denominators have found up to 1174. (denomfind)
500. 7 326,228,178,154,148,132,74 (2q)
675 7 440,308,240,208,200,178,100 (2q)
750. 7 489,342,267,231,222,198,111 (3q)
825 7 538,376,294,254,244,218,122 (2q)
953 7 621,435,339,294,282,252,141 (3q)
1000. 7 652,456,356,308,296,264,148 (4q)
> █
```

コマンドprimefindについて

1. 入力の各行の最初の数値を読み取る。
2. 逐次、入力の各行を読む段階で読み取った数値が、既に読み取った数値の倍数かどうかを判定。そうであれば出力しないし、そうで無ければ、入力行をそのまま出力する。
3. **seq 2,100 | primefind** とすると100以下の素数。
4. このことにより、denomfindで算出した分母の候補について、既に候補となった値の倍数が、全て出力される冗長さを抑制できる。

denomfindの作成時の難点

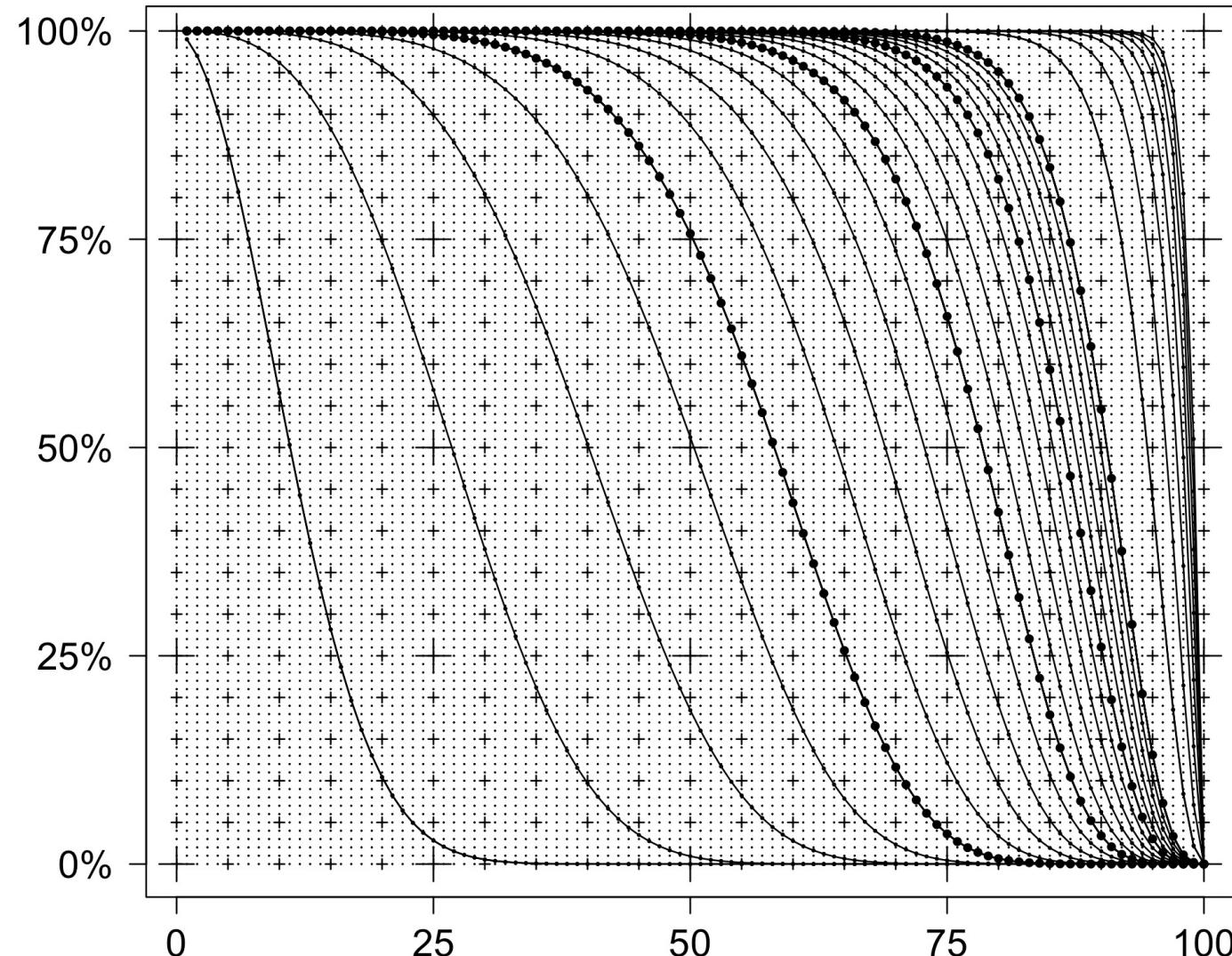
1. 多数の試行錯誤: 出力の形式とオプションの設計
2. 重大なバグを見つけた:
 1. 10%の指定を、近似値0.1と見なし0.05超0.15以下と解釈。
 2. 1.0%の指定を 1%と解釈して区間が**10倍**にしてしまう。
 - 2008年頃に初めてRubyで作ったので慣れていたと思ったが、2018年に作ったPerlのプログラムにあった。
3. 作り込むべき機能が、今後使っているとまだ見つかりそう。



他のプログラムについて

- 2008年頃に一度Rubyで作った。
- primefindはdenomfindと共にApp::denomfindに同梱。
- App::colorplus のcolorplus で colorplus -0により着色を消すことができる。
- App::expandtab のexpandtab でタブ文字の縦揃えが出来る。

小数点以下2桁の割合近似値が k 個あれば
共通分母の推定が分母候補の最小値として求まる確率
についての考察



$$\prod_{i=1}^D \left(1 - (10^{-2}i)^k\right) \quad \text{横軸は } D < 1; k = 1, 2, 3, 4, 5, \dots, 18, 19, 20, 30, 40, \dots, 100 \text{ をプロット。}$$

```

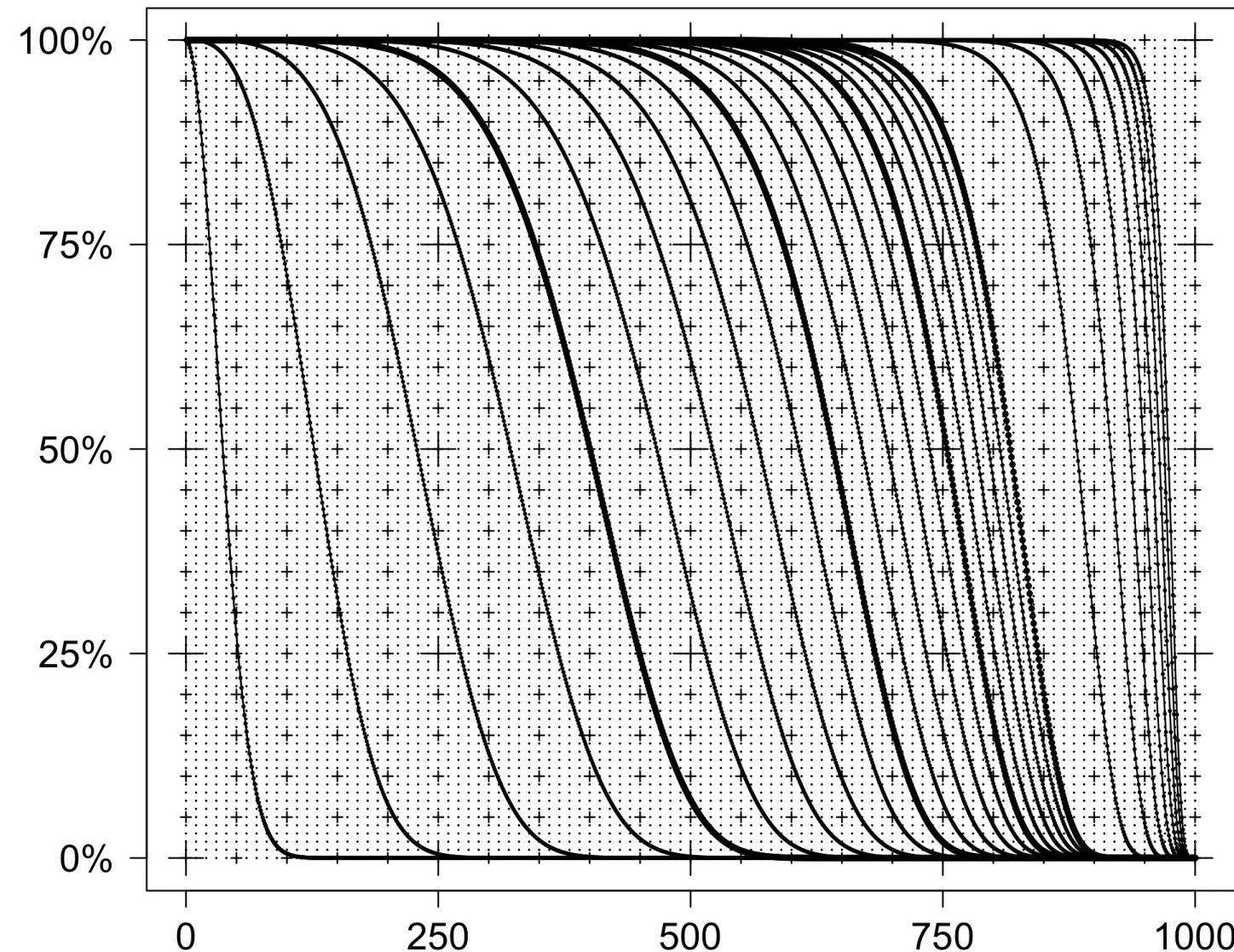
library(matlab)

P <- array(NA,dim=c(100,100))
for (k in 1:100) {
  p <- 1
  for( d in 1:100) { P[k,d] <- p <- p * ( 1- (1e-2*d)^k) }
}

plot ( 1:100, P[1,1:100] , xaxt="n",yaxt="n", xlab="",ylab="",
cex=0.2,type='o')
axis(2,0:4/4,c("0%","25%","50%","75%","100%"),las=1,cex.axis=1.4)
axis(1,0:4*25,cex.axis=1.4)
for(k in 1:20){ points ( 1:100, P[k,] , cex=0.2 , type="o") }
for(k in 1:4*5){ points ( 1:100, P[k,] , cex=0.6 , type="o") }
for(k in 1:10*10){ points ( 1:100, P[k,] , cex=0.2 , type="o") }
points( meshgrid( 0:4*25 , 0:4/4) , pch=3 ,cex=2)
points( meshgrid( 0:20*5 , 0:20/20) , pch=3 ,cex=0.6)
points( meshgrid( 0:100*1 , 0:100/100) , pch=3 ,cex=0.04)

```

小数点以下3桁の割合近似値が k 個あれば
 共通分母の推定が分母候補の最小値として求まる確率
 についての考察



$$\prod_{i=1}^D \left(1 - \left(10^{-3}i \right)^k \right) \quad \text{横軸は } D < 1000; k = 1, 2, 3, 4, 5, \dots, 18, 19, 20, 30, 40, \dots, 100 \text{ をプロット。} \quad 34$$

```

library(matlab)

P <- array(NA,dim=c(100,1000))
for (k in 1:100) {
  p <- 1
  for( d in 1:1000) { P[k,d] <- p <- p * ( 1- (1e-3*d)^k) }
}

plot ( 1:1000, P[1,1:1000] , xaxt="n",yaxt="n", xlab="",ylab="",
cex=0.2,type='o')
axis(2,0:4/4,c("0%","25%","50%","75%","100%"),las=1,cex.axis=1.4)
axis(1,0:4*250,cex.axis=1.4)
for(k in 1:20){ points ( 1:1000, P[k,] , cex=0.2 , type="o") }
for(k in 1:4*5){ points ( 1:1000, P[k,] , cex=0.4 , type="o") }
for(k in 1:10*10){ points ( 1:1000, P[k,] , cex=0.2 , type="o") }
points( meshgrid( 0:4*250 , 0:4/4) , pch=3 ,cex=2)
points( meshgrid( 0:20*50 , 0:20/20) , pch=3 ,cex=0.6)
points( meshgrid( 0:100*10 , 0:100/100) , pch=3 ,cex=0.04)

```