

Mathematica(ver11)+SIR理論 による Covid-19 のデータの分析

§1. 不思議の国の γ

テーマはあくまでも「Mathematicaによるデータ分析の試み」です。Mathematica に対しても Covid-19 に関しても 専門家ではないので、色々間違いがあるかも知れません。 ^_^;
2020年5月11日.(5月19日改定) 生越 茂樹. oh.shigeki@gmail.com

```
In[1]:= ClearAll["Global`*"]
(* 時系列リストdata={v1,v2,...vn}の変域tをbrに従って分割し,dataの区分近似直線関数を作る。関数のリストを作成する関数がfuncs,
傾きのリストを作成する関数がslopes。funcsを組み合わせてdataの折れ線近似を作る関数が multifit.
slopesを組み合わせてdataの傾きの棒グラフを作る関数が multislope。共にtの関数.*)
zip[A_List, B_List] := Partition[Riffle[A, B], 2]
slopes[data_, br_] := Module[{a, b}, Table[
  a /. FindFit[zip[Range[br[[k]], br[[k+1]]], Take[data, {br[[k]], br[[k+1]]}], a x + b, {a, b}, x], {k, 1, Length[br] - 1}]
pieces[data_, br_] := Table[Fit[zip[Range[br[[k]], br[[k+1]]], Take[data, {br[[k]], br[[k+1]]}], {1, x}, x],
  {k, 1, Length[br] - 1}]
multifit[data_, br_] := Which @@ Flatten[Table[{br[[k]] <= x <= br[[k+1]], pieces[data, br][[k]]}, {k, 1, Length[br] - 1}]] /.
  {x -> t}
multislope[data_, br_] := Which @@ Flatten[Table[{br[[k]] <= x <= br[[k+1]], slopes[data, br][[k]]}, {k, 1, Length[br] - 1}]] /.
  {x -> t}
```

■ 0. SIR理論

s = Susceptible (感染可能人数), i = Infected (感染者), r = Recovered (回復者, 死者含む), c = Confirmed Case = $i + r$ とすると
(この r と R_t は全く異なる。また s, i, r, c をプログラム中では $sus, inf, rec, cases$ と書く), SIR理論では,

$$\frac{ds}{dt} = -\beta s * i, \quad \frac{di}{dt} = \beta s * i - \gamma i, \quad \frac{dr}{dt} = \gamma i$$

N を全人口とすると $i + r + s = c + s = N$ だから,

$$\frac{ds}{dt} = -\beta s * i, \quad \frac{dc}{dt} = \beta s * i, \quad \frac{di}{dt} = \beta s * i - \gamma i, \quad \frac{dr}{dt} = \gamma i \quad \dots \text{(#1)}$$

故に $R_t = \frac{\beta \beta}{\gamma}$ と定義すると (これから $R_0 = \frac{N\beta}{\gamma}$, $R_t = R_0 * \frac{s}{N}$ となる),

$$\frac{ds}{d(\gamma t)} = -R_t * i, \quad \frac{dc}{d(\gamma t)} = R_t * i, \quad \frac{di}{d(\gamma t)} = R_t * i - i = (R_t - 1) i, \quad \frac{dr}{d(\gamma t)} = i \quad \dots \text{(#2)}$$

さらに 時間の単位を $1/\gamma$ 日に変えて $\gamma t = t'$ とすると (例えば $\gamma = 0.1$ のときは 10 日ごとの変化を考えると),

$$\frac{ds}{dt'} = -R_t * i, \quad \frac{dc}{dt'} = R_t * i, \quad \frac{di}{dt'} = (R_t - 1) i, \quad \frac{dr}{dt'} = i$$

左の2式から R_t は 感染期間 (回復期間) 毎に, 1 人の感染者が2次感染させる人数 (即ち, 1 人の感染者が2次感染させる人数) となる事がわかる。

(#2 より $R_t = dc / dr$ となることから同じ事が分かる)。

また右から2番めの式から $R_t > 1$ のとき I は増加, $R_t < 1$ のとき I は減少と分かる。

さらに $m_r = \frac{d(\log r)}{dt}$, $m_c = \frac{d\log[c]}{dt}$ と定義すると,

$$m_r = \frac{d(\log r)}{dt} = \frac{1}{r} * \frac{dr}{dt} = \frac{\gamma * i}{r} \quad \therefore \gamma = \frac{r * m_r}{i}, \quad m_c = \frac{d\log[c]}{dt} = \frac{1}{c} * \frac{dc}{dt} \quad \therefore \frac{dc}{dt} = c * m_c. \quad \dots \text{(#3)}$$

故に

$$R_t = \frac{\beta \beta}{\gamma} \left(= \frac{1}{i * \gamma} \frac{dc}{dt} \right) = \frac{c * m_c}{i * \gamma} = \frac{c * m_c}{r * m_r} \left(= \frac{dc}{dr} \right) \quad \dots \text{(#4)}$$

■ 1. Johns Hopkins 大学 (GitHub) から Download

Johns Hopkinsからのデータでは recovered は 死者を含まないが, SIR理論では 死者も含む。以下
「r」または「rec」と書く SIR理論の回復者の意味とする。日本語で「回復者」と書く, 死者を含まないとする。

ConfirmedCaseは感染者の積算データなので、infected= active case(感染者.以下「inf」)は、それから回復者と死者を引いたものになる。さらに(回復者+死者)が50人になった日から改めて数える。以下、day1→recが50人以上となった日のindex. day1→ inf<10 または最新データ獲得日のindex. inf0,rec0,deaths0,cases0,days0→全てのデータ. inf,rec,deaths,cases,days→day1からのデータ、period→day1からday1までの日数。

```

In[7]:= casesGlobal = Import[
  "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/
    time_series_covid19_confirmed_global.csv"];
deathsGlobal = Import[
  "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/
    time_series_covid19_deaths_global.csv"];
recGlobal = Import[
  "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/
    time_series_covid19_recovered_global.csv"];
countries = Drop[casesGlobal[[All, 2]], 1];

nation = "Japan"; (*上のリストから. United States は US. UK は United Kingdom.*)

cases0 = casesFromJH = Plus @@ (Drop[#, 4] &) /@ Select[casesGlobal, Part[#, 2] == nation &];
recFromJH = Plus @@ (Drop[#, 4] &) /@ Select[recGlobal, Part[#, 2] == nation &];
deaths0 = deathsFromJH = Plus @@ (Drop[#, 4] &) /@ Select[deathsGlobal, Part[#, 2] == nation &];
rec0 = recFromJH + deathsFromJH;
inf0 = cases0 - rec0;

day1 = FirstPosition[ rec0 - 50, _?Positive][[1]]; (*rec>50からカウント*)
dayInfLess10 = Flatten[Position[ inf0 - 10, _?Negative]];
day1 = Min[Select[dayInfLess10, # > day1 &] - 1, Length[cases0]]; (*inf<10になると中止*)
datesGlobal = Drop[casesGlobal[[1]], 4];
period = day1 - day1 + 1;
inf = Take[ inf0, {day1, day1}];
rec = Take[ rec0, {day1, day1}];
cases = Take[ cases0, {day1, day1}];
deaths = Take[ deaths0, {day1, day1}];
dates = Take[datesGlobal, {day1, day1}];

```

■ 2. Log(rec)の傾き m_r を求める.

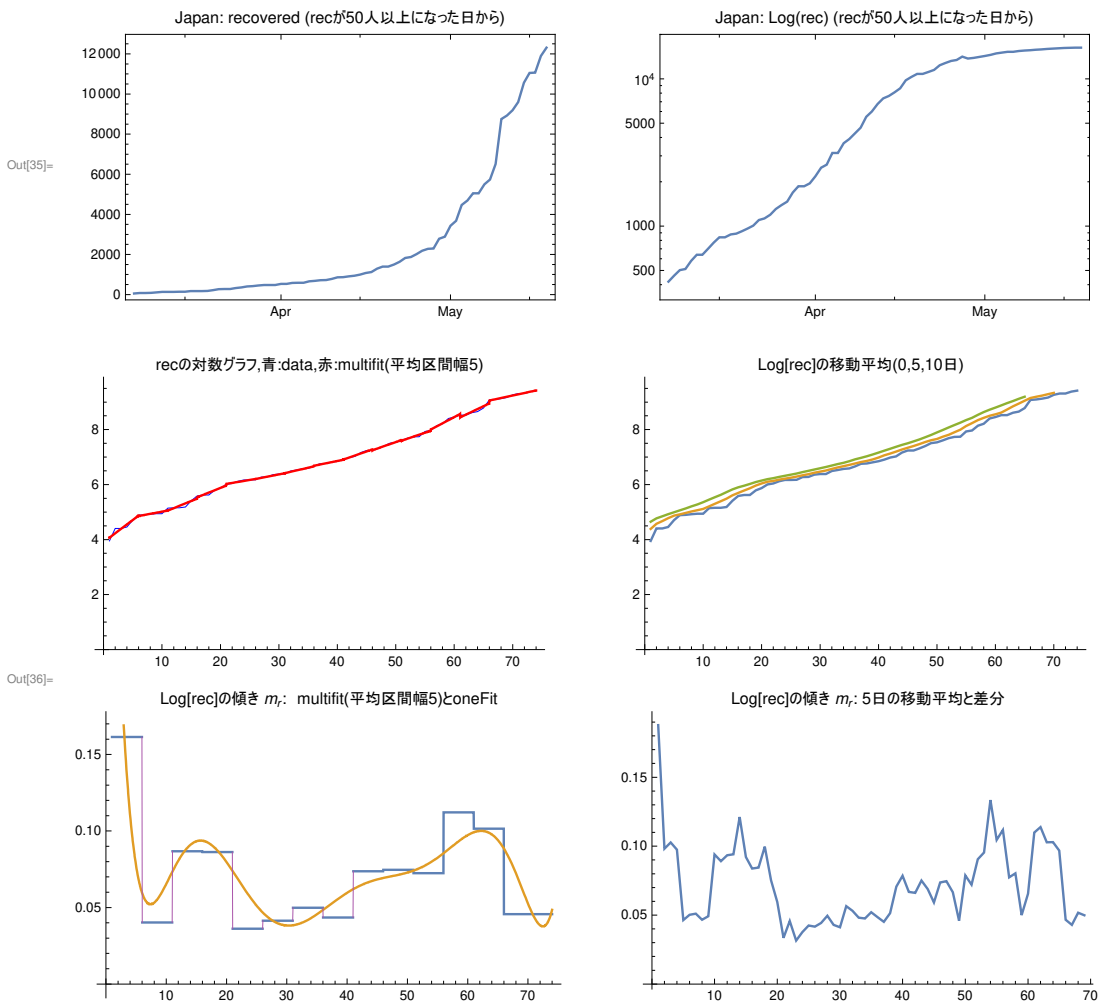
「移動平均+差分を使う方法」と「全体のFit関数の微分(oneFit)」と「下のbreaksに従って期間を分割してFitを使う方法(multifit)」を比較する。また breaksは、Default では「始点はt=1, 幅5, 最後のみ少し長め」に設定。しかし グラフを目で確認しながら設定可能。(但し最初は1,最後はperiodにする)

```

In[27]:= haba = 5; (*←幅を設定*)
breaks = Table[1 + haba k, {k, 0, (period - 1) / haba - 1}] // Append[period];
(*←グラフを目で見て直線部分を見つけると、更に適切*)
mhaba = Round[Mean[Take[RotateLeft[breaks] - breaks, {1, Length[breaks] - 1}]];
Print[Style["breaks (解析期間の分割) =" <> ToString[breaks], Larger, Blue]]

logRmultifit[t_] = multifit[Log[rec], breaks]; (*Log(rec)の折れ線近似*)
mr[t_] = multislope[Log[rec], breaks]; (* $m_r$ の棒グラフ*)
mrByOneFit = D[Fit[Log[rec], Table[t^k, {k, 0, 10}], t], t]; (*全体にFitを使用し微分-onefit*)
mrByMovingAverage = Differences[MovingAverage[Log[rec], 5]];
GraphicsRow[{DateListPlot[TimeSeries[rec, {dates}], PlotLabel → nation <> ": recovered (recが50人以上になった日から)",
  DateListLogPlot[TimeSeries[cases, {dates}], PlotLabel → nation <> ": Log(rec) (recが50人以上になった日から)",
  ImageSize → 700]} // Quiet
GraphicsGrid[{{Show[{ListLinePlot[Log[rec], PlotStyle → {Thin, Blue}],
  Plot[logRmultifit[t], {t, 1, period}, PlotStyle → Red],
  PlotLabel → "recの対数グラフ" <> ", 青:data, 赤:multifit (平均区間幅" <> ToString[mhaba] <> ")"], ListLinePlot[
  {Log[rec], MovingAverage[Log[rec], 5], MovingAverage[Log[rec], 10]}, PlotLabel → "Log[rec]の移動平均 (0,5,10日)"}],
  {Plot[{Tooltip[mr[t], "multifit"], Tooltip[Evaluate[mrByOneFit], "onefit"]],
  {t, 1, period}, AxesOrigin → {0, 0}, ExclusionsStyle → Lighter[Purple],
  PlotLabel → "Log[rec]の傾き  $m_r$ :" <> " multifit (平均区間幅" <> ToString[mhaba] <> ")とoneFit"},
  ListLinePlot[mrByMovingAverage, PlotLabel → "Log[rec]の傾き  $m_r$ : 5日の移動平均と差分"}]}, ImageSize → 700}
breaks (解析期間の分割) = {1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 74}

```



3つのグラフの傾向は同じだが、移動平均は、やや変動が大きく、かつ変域が左にずれる。onefit と multifit を比べると onefit の方は両端が大きくハズれる。次章で扱う log(cases)の傾き m_c と違い はっきり言ってどちらを採用してもよいのだが m_c に合わせて「階段関数 multifit」を m_r に採用したい。即ち $m_r = mr[t]$ とする。

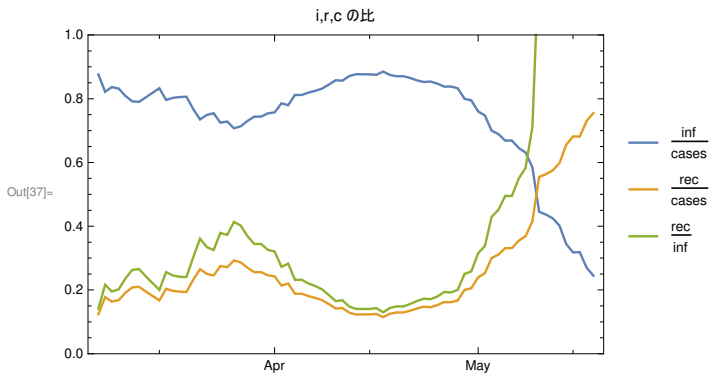
■ 3. m_r から γ を求める.

SIR理論では

$$ds/dt = -\beta s * i, \quad di/dt = \beta s * i - \gamma i, \quad dr/dt = \gamma i \quad (\beta, \gamma \text{は定数})$$

一番右の式から $\frac{d(\log r)}{dt} = \frac{1}{r} \cdot \frac{dr}{dt} = \frac{\gamma * i}{r}$. よって $y = \log r$ のグラフの傾きを m_r とすると, $\gamma = m_r \cdot \frac{r}{i}$

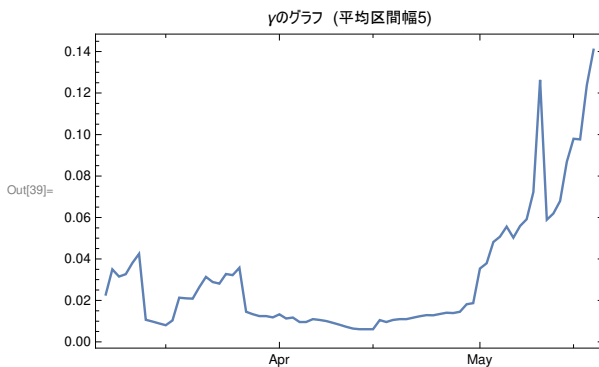
```
In[37]= DateListPlot[TimeSeries[#, {dates}] & /@ {inf / cases, rec / cases, rec / inf},
PlotLegends -> {"inf / cases", "rec / cases", "rec / inf"}, PlotRange -> {0, 1}, PlotLabel -> "i, r, c の比"]
```



$\gamma = m_r * r / i$ だから, これから γ のグラフが描ける.

r/i は短い期間では一定としてよいはずなのだが, 最近の急上昇を見るとそうも言えない. したがって r/i は平均で考えずに時間 t における値で考える.

```
In[38]= \gammaTable = Table[mr[t], {t, 1, period}] * rec / inf; (*\gamma=m_r*r/i. 区分関数*)
DateListPlot[TimeSeries[\gammaTable, {dates}],
PlotLabel -> "\gammaのグラフ" <> " (平均区間幅" <> ToString[mhaba] <> ") ", AxesLabel -> "\gamma"]
```



本来は γ は「回復日数の逆数 = 感染期間の逆数」なので, 定数と考えられる (統計のとり方の違いで国には依るだろう) しかし上のグラフを見ると, 定数と見ることは難しい. (5/11では定数と見ることができたが 5/19にはもはや無理.) とりあえず直近を除いて平均をとってみると,

```
In[40]= \gamma = TrimmedMean[\gammaTable, {0.2, 0.2}] // N;
Print[Style["単純平均によると {\gamma, 回復(感染)日数}={", <> ToString[\gamma] <> ",", <> ToString[1/\gamma] <> "日}", Larger, Blue]]
単純平均によると {\gamma, 回復(感染)日数}={0.0213305, 46.8813日}
```

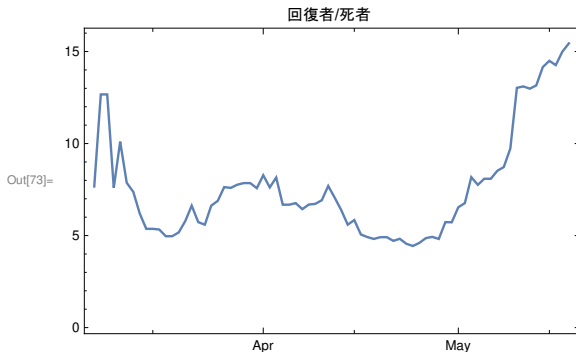
上の回復(感染)日数は長過ぎる. コロナに2ヶ月も罹っているとは考えられない. 一方, 直近の値では $\gamma=0.13$ (回復日数=8日) すごい違いがある. この様に r の変化だけでは γ の値を決めようがないが, 次章で微分方程式と組み合わせて γ と R_t を一度に求めてみたい.

[注]この章を最初に書いた5/11時点では γ の急上昇がなかったため $\gamma=0.016$ だった. そのとき「実際には r が統計データより大きい可能性が高い. 実際, 回復者のデータが何日も更新されない国もあり worldoddata.org では回復者のデータは置いてないので十分可能性がある」と書いてその理由をもう一つ上げた. 以下は $\gamma=0.016$ (5/11時点)に対して, 書いた文である.

■ 4. r は統計データより多いはず (致死率との関係)

($\gamma=0.016$ (5/11時点)に対して、書いた文である。回復者は死者を含めないとする。但しrec は 死者を含めるとする。)

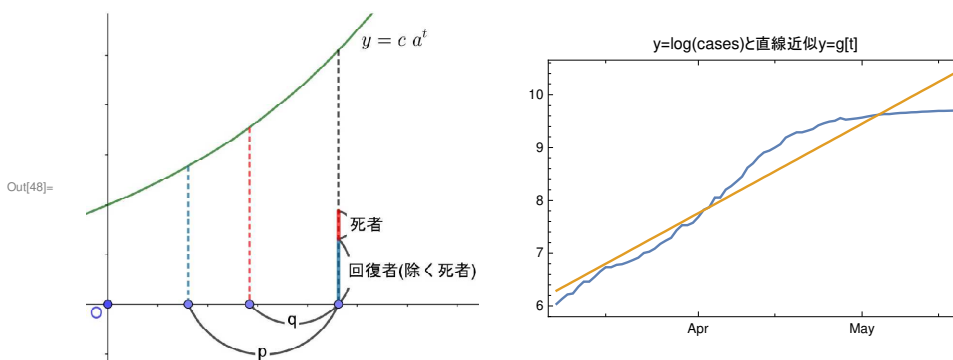
```
In[73]:= DateListPlot[TimeSeries[{{(rec - deaths) / deaths}, {dates}], PlotLabel -> "回復者/死者 "]
TrimmedMean[N[(rec - deaths) / deaths], {0.2, 0.2}];
Print[Style["回復者/死者=" <> ToString[%], Blue, Larger]]
```



回復者/死者=6.85689

上を見ると回復者は死者の6.5倍しかないが、日本の致死率は2%ぐらいである。ところが回復者が死者の6.5倍なら、致死率ももっと高いはずである。(1/7.5=13%) これは回復するのに掛かる平均日数をp日、死亡するのに掛かる平均日数をqとしたとき($p=\text{回復日数}=1/\gamma$)、十分に $p>q$ なら説明できる。

```
In[45]:= sol = FindFit[Log[cases], m t + n, {m, n}, t];
g[t_] = (m t + n) /. sol;
a = Exp[m] /. sol;
GraphicsRow[{Import["http://mixedmoss.com/miscellaneous/MathematicaSIRModel/deathRatio.jpg"],
DateListPlot[TimeSeries[#, {dates}] & /@ {Log[cases], Table[g[t], {t, 1, period}]},
PlotLabel -> "y=log(cases)と直線近似y=g[t]"},
ImageSize -> 600]
Print[Style["log(cases)=" <> ToString[g[t]] <> ", cases=c*a^t (a=" <> ToString[a] <> ")", Blue, Larger]]
```



$\log(\text{cases}) = 6.23409 + 0.0564484 t$, $\text{cases} = c * a^t$ ($a = 1.05807$)

Confirmed cases は基本的に指数関数なので $\text{cases} = c * a^t = f(t)$ (a, c 定数) とおける。(5/11時点では)
 $p > q$ のとき、回復者/死者が6.5、死亡率が2%となることから、

$$6.5 = \text{回復者/死者} = \frac{f(t-p) * 0.98}{f(t-q) * 0.02} = 49 * a^{q-p} \quad \times$$

$$\text{故に } p - q = \log_a(49 / 6.5)$$

\times
casesのデータからaを概算で求めると $a = 1.07$ で

$$p = 1 / \gamma = 1 / 0.0166 = 60 \text{ 日,}$$

$$q = p - \log_{1.07}(49 / 6.5) = 30 \text{ 日}$$

となる。これはおかしい。故に γ は0.0166より大きくないといけない。

[結論] 基本的な考え方は今でも正しいと思っているが, 5 / 19 時点ではもはや「 $\text{cases} = c * a^t$ 」が成り立たない.
(右上図で大きく直線からずれている) したがって上の式そのものは改良しないと成り立たない.

何れにせよ γ の値は 0.016 より大きいことは間違いない. 正確な値は微分方程式を解き, 総合的判断も入れて次章で求めたい.

■ 5. 次章のため「nation, dates , breaks ,cases,rec,deaths」を Export

```
In[50]:= Export["nation.txt", nation];  
Export["numbers.csv", {dates, breaks, cases, rec, deaths}];
```