

2014 年度統計数理研究所共同研究集会

データ解析環境 R の整備と利用

# による数理疫学

谷村 晋


arumiant@gmail.com

兵庫医科大学公衆衛生学講座

2014 年 11 月 29 日



# 本日の内容 I

- 1 数理疫学とは
- 2 数理疫学の Hellow World!
- 3  $R_0$  と 
- 4 数理モデルへの批判



# 疫学と数理疫学

- **疫学**とは、特定の集団における健康に関連する状況あるいは事象の、分布あるいは規定因子に関する研究。また、健康問題を制御するために疫学を応用すること（国際疫学学会）
- 疫学研究では、観察研究（記述疫学、分析疫学）や実験研究（介入研究、フィールド試験）などが用いられるのが一般的であるが、数理モデルを使う研究もある→**数理疫学**

# 数理疫学


- 数理モデルを使って、疫学的健康事象に検討を加えて、何らかの解釈を得るアプローチ→得られた研究成果は疾病対策の科学的根拠となる
- 主に病因と宿主の相互作用を説明するために、生物学的、行動科学的、免疫学的、人口学的、医学的視点を数理モデル化する（**感染症の数理モデル**）
- 応用事例
  - どのくらいの割合の人にワクチンを接種すれば感染の流行を抑止できるか→麻疹の流行を抑えるには94%以上のワクチン接種率が必要
  - 地域封鎖に必要な半径はどの程度か→新型インフルエンザでは10kmが必要（通勤電車で高感染の場合は13km）

# 数理疫学の過去と現在

- 1920年代に Reed-Frost モデル（連鎖二項モデル）や Kermack-McKendrick モデル（微分方程式）
- 和文雑誌『理論疫学研究』（1954–1989年、阪本州弘先生，兵庫医大）
- 1980年代以降、新興感染症、再興感染症の脅威により、感染症の数理モデルの研究が欧米で加速する
- 西浦博先生（当時：長崎大学熱帯医学研究所）が世話役となり、「感染症理論疫学研究大会」が開催（2006年）
- 稲葉寿著『感染症の数理モデル』（2008年）



# 本日の内容 I

- 1 数理疫学とは
- 2 数理疫学の Hellow World!
- 3  $R_0$  と 
- 4 数理モデルへの批判



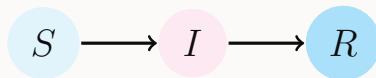
# SIR モデル

- 数理疫学の講義で最初に説明する最も単純なモデル → SIR モデル
- Kermack と McKendrick (1927) が提案した古典的モデル
- 局地的な感染症流行によく当てはまる
- 数多くある派生モデルの原点



# SIRモデル

3つの人口集団を想定する



- $S$ : 感受性のある集団
- $I$ : 現在感染しており、感染性を有している集団
- $R$ : 免疫を持っており、以降は感染しない集団

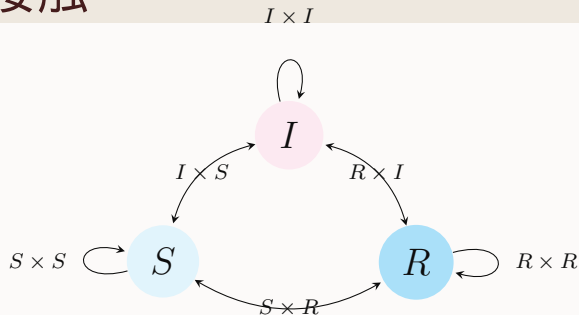


# 単純化のための前提

- 1 サイズが  $N$  の閉鎖集団
- 2 感染率や免疫獲得（または死亡）率は不変
- 3 出生も死亡も無視できる程度に短期間
- 4  $S$  も  $I$  も  $R$  も十分に混ざって均等化している人口集団
- 5 潜伏期間はなし、感染期間と発症期間のずれもなし



# SIRの接触



- 感受性を持つ人と感受性を持つ人が会う ( $S \times S$ )
- 感受性を持つ人と感染している人が会う ( $S \times I$ ) ← 感染！
- 感受性を持つ人と免疫を獲得した人が会う ( $S \times R$ )
- 感染している人と感染している人が会う ( $I \times I$ )
- 感染している人と免疫を獲得した人が会う ( $I \times R$ )
- 免疫を獲得した人と免疫を獲得した人が会う ( $R \times R$ )



# SIR のモデル式

## 常微分方程式システム

$$\begin{aligned}
 \frac{d}{dt}S(t) &= -\beta\kappa S(t)I(t) \\
 \frac{d}{dt}I(t) &= \beta\kappa S(t)I(t) - \frac{I(t)}{D} \\
 \frac{d}{dt}R(t) &= \frac{I(t)}{D}
 \end{aligned} \tag{1}$$

$S(t)$ ,  $I(t)$ ,  $R(t)$  はそれぞれの  $t$  時点の人口割合。 $\beta$  は 1 回の接触あたりの感染確率（接触感染確率）、 $\kappa$  はある時間あたり 1 人の人間が集団内で平均何回接触するのか、 $D$  は感染期間



# SIR のモデル式

モデル式 (1) を  $\mathbb{R}$  でコーディングしやすいように書き換える

$$\begin{aligned}
 S_{t+1} &= S_t - \beta\kappa S_t I_t / N \\
 I_{t+1} &= I_t + \beta\kappa S_t I_t / N - I_t / D \\
 R_{t+1} &= R_t + I_t / D
 \end{aligned} \tag{2}$$

$S_t, I_t, R_t$  はそれぞれの  $t$  時点の人口。 $\beta$  は 1 回の接触あたりの感染確率（接触感染確率）、 $\kappa$  はある時間あたり 1 人の人間が集団内で平均何回接触するのか、 $N$  は総人口、 $D$  は感染期間



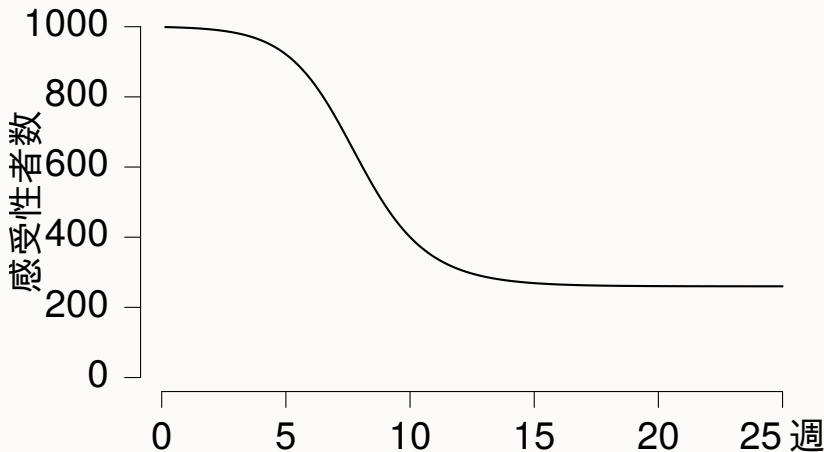
## SIRモデルのRコード

```

plotSIR <- function(n = 1000, b = 0.15, k = 12/7, D = 1 *
  7, day = 7 * 25) {
  S <- I <- R <- integer(day)
  I[1] <- 1
  S[1] <- n - I[1]
  for (i in 1:(day - 1)) {
    S[i + 1] <- S[i] - b * k * S[i] * I[i]/n
    I[i + 1] <- I[i] + b * k * S[i] * I[i]/n - I[i]/D
    R[i + 1] <- R[i] + I[i]/D
  }
  res <- data.frame(S, I, R)
  matplot(res, type = "l", lwd = 2, las = 1, ylab = "人数",
    xaxt = "n", bty = "n")
  xw <- (0:5) * 5
  xw[length(xw)] <- paste0(xw[length(xw)],
    "週")
  axis(side = 1, at = (0:5) * 7 * 5, labels = xw)
  return(invisible(res))
}

```

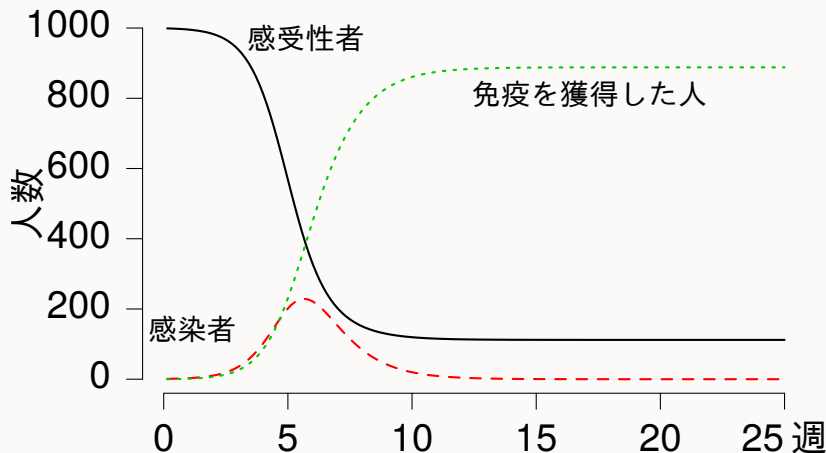
# Sの変化



- 人口千人の閉じられた集団で想像上の流行が起こった場合 ( $\beta = 0.15$ ,  $\kappa = 12/7$ ,  $D = 1 \times 7$ ) の感受性者数の推移 (流行ゼロ週に1人の患者が発生した場合)



# SIR の変化

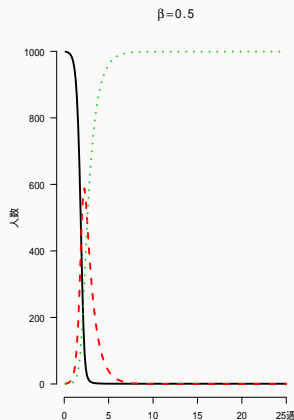
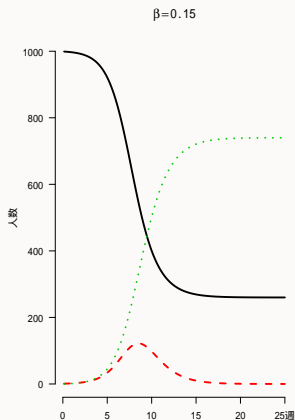
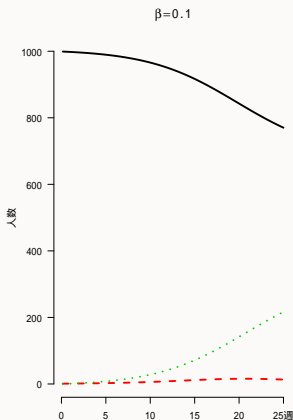


- 想像上の流行が起こった場合 ( $\beta = 0.2, \kappa = 12/7, D = 1 \times 7, n = 1000$ ) の感染者数などの推移



# 感染力の影響

$\beta$  を変えてみる

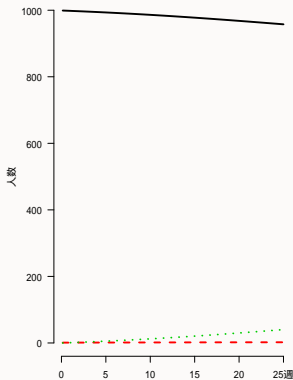




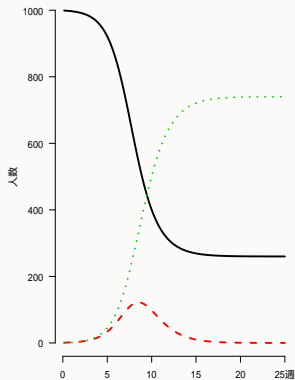
# 接触回数の影響

$\kappa$  を変えてみる

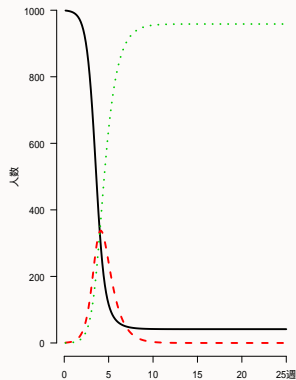
$\kappa=7/7$



$\kappa=12/7$



$\kappa=21/7$

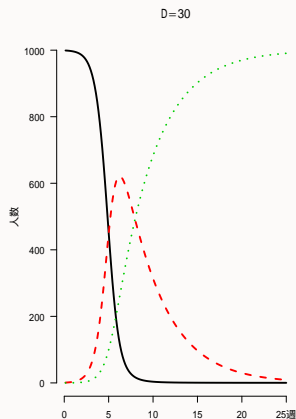
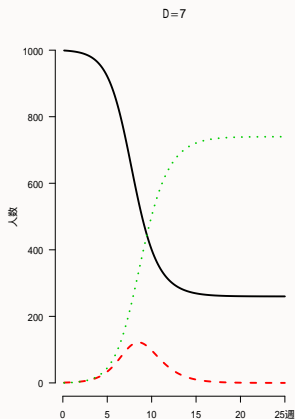
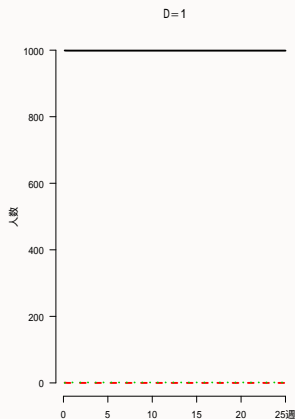


学級閉鎖は感染症流行を抑制する



# 感染期間の影響

$D$  を変えてみる

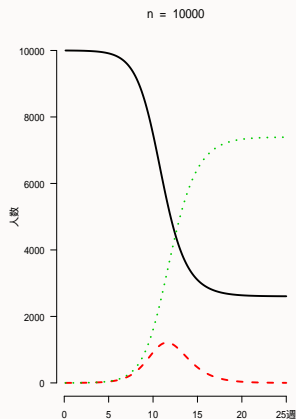
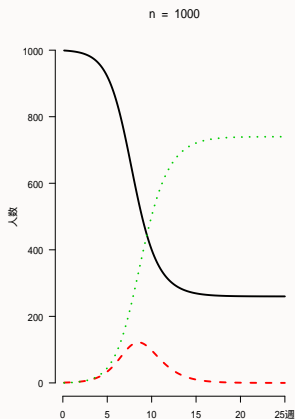
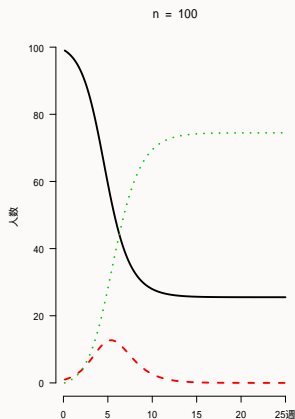


感染期間が長いと大流行



# 人口サイズの影響

$N$  を変えてみる



人口が多いと罹患人口が増える→都市部で流行すると深刻



# 侵入可能な条件

モデル式 (1) の  $I(t)$  について、流行のごく初期では  $S(t) \approx 1$  と見なせるので、

$$\frac{d}{dt}I(t) = \beta\kappa S(t)I(t) - \frac{I(t)}{D} = \left(\beta\kappa - \frac{1}{D}\right)I(t)$$

従って、人口集団に感染症が侵入するためには、 $\beta\kappa - \frac{1}{D} > 0$  である必要がある



$R_0$ 

rought

$\beta\kappa N - \frac{1}{D} > 0$  から  $R_0$  を定義する

$$R_0 = \beta\kappa D > 1$$

- $R_0 > 1$  ならアウトブレイク発生、 $R_0 < 1$  なら流行せずに自然消滅
- 感染症の基本再生産数（比）→一人の感染者から平均で何人の感染者が新たに感染するか
- $R_0$  の考え方は古くからある (G. McDonald, 1952) が、感染症流行対策において、重要な疫学指標の1つ→罹患者数予測、ワクチンの効果による感染防御者数の予測など
- $R_0$  は、感染者の感受性、病原体の感染力、感染可能な人口、接触頻度・様式などに影響を受ける




# 主要な感染症の $R_0$

疾患	伝播様式	$R_0$
麻疹 (はしか)	空気感染	12-18
百日咳	空気・飛沫感染	12-17
ジフテリア	飛沫感染	6-7
ポリオ	糞口感染	5-7
風疹	空気・飛沫感染	5-7
HIV/AIDS	性的感染	2-5
SARS	空気・飛沫感染	2-5
インフルエンザ (スペイン風邪)	空気・飛沫感染	2-3
エボラ (2014 アウトブレイク)	接触感染	1.5-2.5

● ただし、 $R_0$  は時代、社会、国、病原体などによって異なる

# 本日の内容 I

- 1 数理疫学とは
- 2 数理疫学の Hellow World!
- 3  $R_0$  と 
- 4 数理モデルへの批判




# R<sub>0</sub> を計算する パッケージ

**R0 パッケージ**  $R_0$  や  $R(t)$  の各種推定法を実装したパッケージ

**surveillance パッケージ** 感染症流行監視システムの総合パッケージ<sup>1</sup>。  $R_0$  推定もできる

その他の関連パッケージ

**simecol パッケージ** 数理生態学モデルを  に実装するパッケージ (SIR モデルも含む)

**outbreaker パッケージ** 病原体の遺伝子配列情報を用いてアウトブレイクを再現するパッケージ →  $R_0$  を引数に使う

<sup>1</sup>オンラインヘルプの PDF は 291 ページ



# R0 パッケージ

- 疾病流行から  $R_0$  と  $R(t)$  (real-time reproduction number) を推定する
- 著者は、インフルエンザの数理疫学で実績のある PY Boëlle 教授 (フランス国立保健医療研究所) と彼が指導する大学院生である T Obadia 氏



写真

PY Boëlle 教授



写真

T Obadia 氏

- C 言語や FORTRAN は使わずに、R だけを用いて書いている。MASS パッケージだけに依存。



# スペイン風邪

- 第一次世界大戦中である 1918 年から 1920 年にかけて H1N1 型の新型インフルエンザが全世界に流行
- 世界人口 18 億人のうち、5–10 億人が感染
- 世界人口の 2–4% にあたる 4000–8000 万人が死亡したと推定される → 第一次世界大戦の全戦没者 1000 万人の数倍

写真

- 第一次世界大戦後のフランスにおける米軍キャンプのインフルエンザ病棟
- アメリカ軍の全戦没者の 80% (99200 人) は、このスペインかぜの感染によるものとも言われる

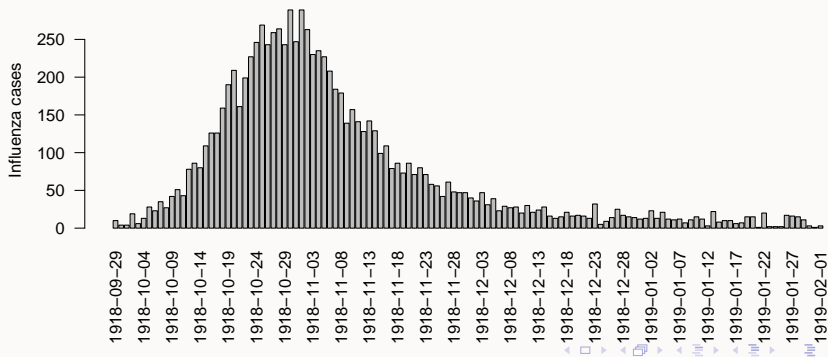
# 日本の被害

- 日本ではスペインかぜで 38 万人が死亡したとされてきたが、最近の再調査では、45 万人以上の方が死亡
- 当時の東京日日新聞の見出し
  - 鉄道員の欠勤が多く、輸送に不便が出る
  - 死者の 4 人に 1 人がインフルエンザ
  - 地方火葬場へ送るため停車場に死体の山
  - 農村で全戸全滅

# ドイツにおけるスペイン風邪

ドイツ帝国・プロセイン地方 (計 8911 人)

- > `library(R0)`
- > `## Outbreak during 1918 influenza pandemic in Germany`
- > `data(Germany.1918)`



# 世代時間分布の計算

```
> mGT <- generation.time("gamma", c(2.6, 1))
> str(mGT)
```

List of 4

```
$ GT : num [1:9] 0 0.119 0.392 0.315 0.129 ...
$ time: int [1:9] 0 1 2 3 4 5 6 7 8
$ mean: num 2.6
$ sd : num 1.04
- attr(*, "class")= chr "R0.GT"
```

文献に基づく平均と標準偏差を `generation.time()` に与えて、`R0.GT` クラスオブジェクトを得る



# 当てはめるデータの範囲

- $R_0$  はごく初期の再生産数
- 指数的成長曲線で当てはめる疫学流行曲線の範囲を決めておく必要がある
- 流行開始日からピーク日までの間を用いる<sup>2</sup>→**今回は32日**

<sup>2</sup>厳密に行う場合は適合度が最大になる日数を求める

# $R_0$ の計算法

## ● Exponential growth 法 (EG 法)

- 初期の感染人口は指数関数的成長法則で増加する
- ポアソン回帰モデルを当てはめて指数的成長率  $r$  を得る
- 世代時間分布関数  $M$  から  $R$  を求める

$$R = \frac{1}{M(-r)}$$

## ● 最尤度推定法 (ML 法)

- 時間単位ごとの罹患数  $N_0, N_1, \dots, N_T$  及び世代時間分布  $w$  から、次の対数尤度関数を用いる

$$LL(R) = \sum_{t=1}^T \log \left( \frac{e^{-\mu_t} \mu_t^{N_t}}{N_t!} \right), \quad \text{ただし、} \mu_t = R \sum_{i=1}^t N_{t-1} w_i$$

## EG 法

```
> estR0 <- estimate.R(Germany.1918, mGT, begin = 1, end = 32,
+                      methods = "EG")
> attributes(estR0)
```

```
$names
[1] "epid"      "GT"        "begin"
[4] "end"       "estimates"
```

```
$class
[1] "R0.sR"
```

```
> estR0
```

```
Reproduction number estimate using Exponential Growth method.
R : 1.343115[ 1.32636 , 1.360239 ]
```





# EG 法

## plotfit メソッド

```
> isS4("R0.sR")
```

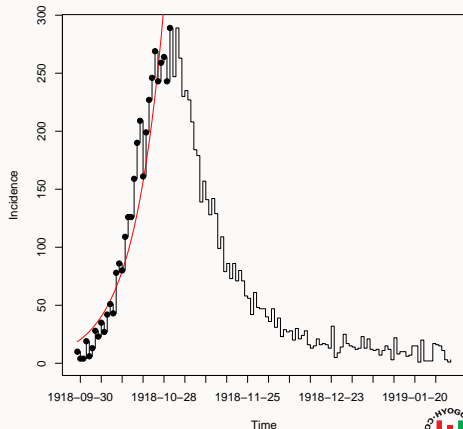
```
[1] FALSE
```

```
> methods(class = "R0.sR")
```

```
[1] plot.R0.sR* plotfit.R0.sR* print.R0.sR*
```

Non-visible functions are asterisked

Epidemic curve & model ( Exponential Growth )



# ML法

```
> estR0 <- estimate.R(Germany.1918, mGT, begin = 1, end = 32,  
+                       methods = "ML")  
> attributes(estR0)
```

```
$names  
[1] "epid"      "GT"        "begin"  
[4] "end"       "estimates"
```

```
$class  
[1] "R0.sR"
```

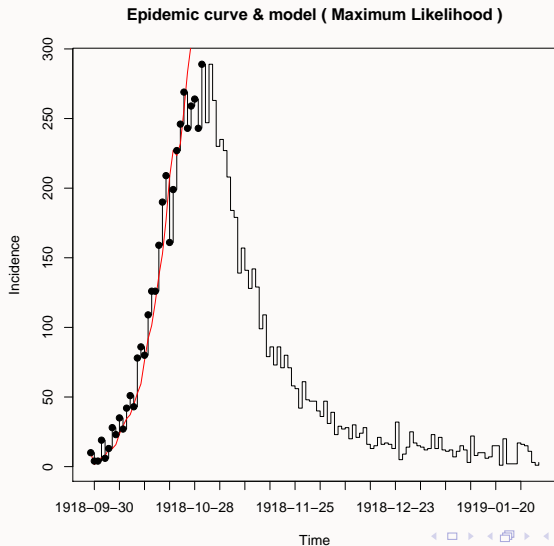
```
> estR0
```

```
Reproduction number estimate using Maximum Likelihood method.  
R : 1.214487[ 1.161099 , 1.269497 ]
```



## ML 法

plotfit メソッド



## 2つの計算結果を比較

```
> estR0 <- estimate.R(Germany.1918, mGT, begin = 1, end = 32,  
+                      methods = c("EG", "ML"))  
> estR0
```


```
Reproduction number estimate using Exponential Growth method.  
R : 1.343115[ 1.32636 , 1.360239 ]
```

```
Reproduction number estimate using Maximum Likelihood method.  
R : 1.214487[ 1.161099 , 1.269497 ]
```

```
>
```



# 本日の内容 I

- 1 数理疫学とは
- 2 数理疫学の Hellow World!
- 3  $R_0$  と 
- 4 数理モデルへの批判






# 感染症の数理モデルへの批判

- 均一な人口集団なんてありえない→性・年齢など人的属性による感受性の違いを無視
- 社会の接触パターンが均一なんてありえない→サブグループ（家族、学校など）内外で異なる、発病すれば病臥して外出しない
- 感染経路が1つという前提がおかしい、感染経路によって感染確率はそれぞれ異なる
- 感染確率、接触回数、感染期間で  $R_0$  が決まり、必ず同じ流行曲線が再現されるはずであるが、最初の感染者が誰にも出会わずに感染期間が終わったら流行しない事象と矛盾する (deterministic model)  
→ stochastic model に発展



# おわりに

- のプラットフォームに既存の各種手法を網羅的にコレクションして、実装されるのは、エンドユーザにとって大変ありがたい（新しい手法を提唱する論文の著者は必ずしもコードを書かない）
- R0パッケージは、提案されている  $R_0$  計算法を網羅していた
- 一方、疫学全体を見ると、いろいろな人が好き勝手にパッケージを書いて、独自のクラスを定義したりして混乱中（疫学の守備範囲が広すぎることも原因）