統計モデルによる
災害のリスク評価

中央大学・理工学部　鎌倉稔成

リスクとは

$$\sum（損失）\times 頻度$$

事象生起とリスク評価の統計的意味

- 事故の生起による，死亡者数の問題を考えてみよう.
  - １年間で平均５回起こる事故があるとき，その１回の事故で平均２人の死亡者があるという.
    １年間で何人の死亡者が見積もられるか.

- １人死亡による補償料として１億円かかるものとすれば，年間予算として
１×10=10億円を用意すればよい？

5件／年　×　2人／件　＝　10人／年

確率モデルの必要性

- 問題を定式化すれば，i回目の事故における死亡者数をXi，年間事故件数をNとすれば，求める年間死亡者数Yは

$$Y = \sum_{i=1}^{N} X_i$$

Wald's Identity

$$E(Y) = E\left[\sum_{i=1}^{N} X_i\right] = E\left[E\left[\sum_{i=1}^{N} X_i \middle| N\right]\right] = E\left[NE(X_1)\right] = E(N)E(X_1).$$

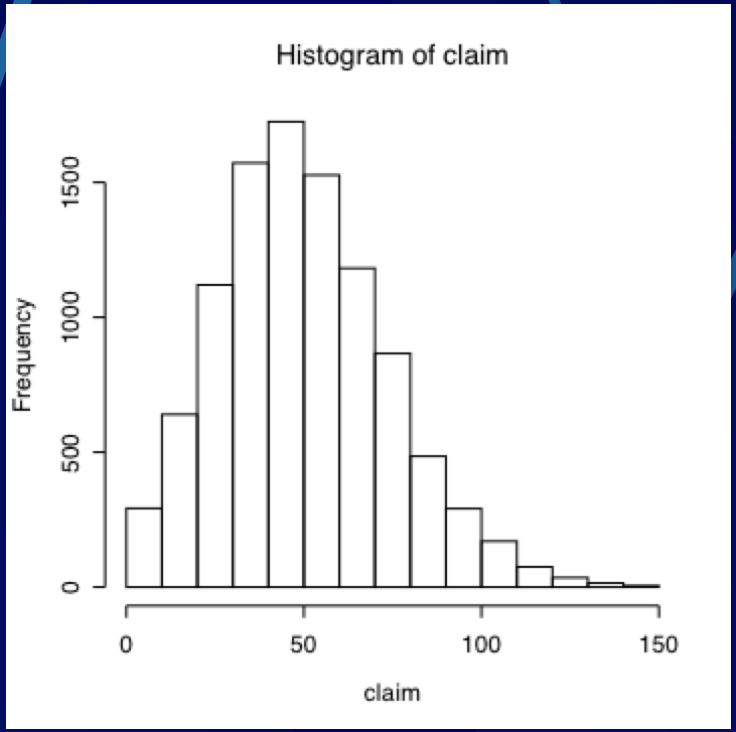
ポアソンモデル

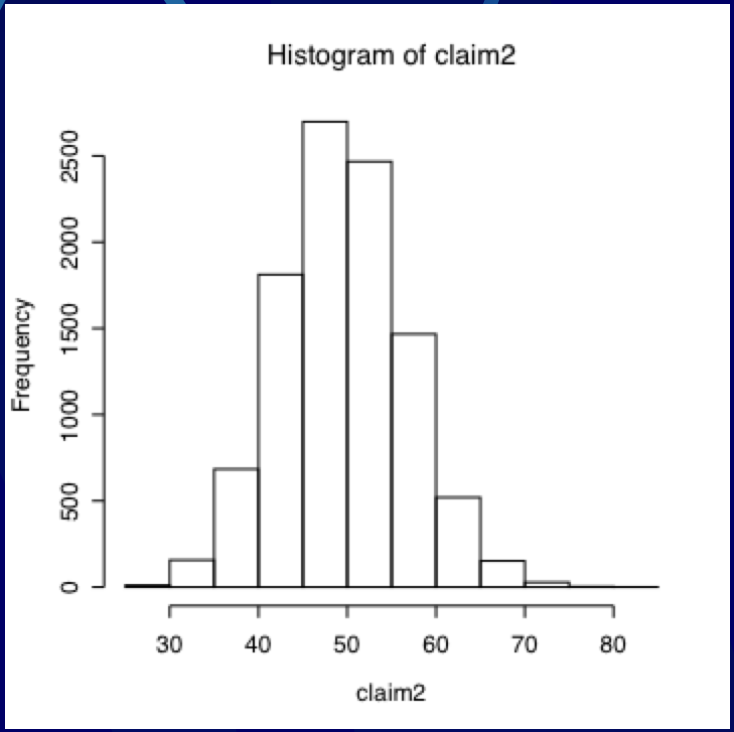
分散については,

$$Var(Y) = E\left[Var(Y|N)\right] + Var\left[E(Y|N)\right].$$

$$\begin{aligned} Var(Y) &= E\left[NVar(X_1)\right] + Var\left[NE(X_1)\right] \\ &= E\left[N\lambda_2\right] + Var\left[N\lambda_2\right] \\ &= \lambda_2 E(N) + \lambda_2^2 Var(N) \\ &= \lambda_1\lambda_2 + \lambda_1\lambda_2^2. \end{aligned}$$

死亡者数の分布





quantile(claim,0.95)=93 quantile(claim2,0.95)=62

事象の生起確率のモデルと検定

- トレンドアナリシスとしての統計モデル
  cumulative mean function

$$M(t) = E\{N(t)\}$$

$$\lambda(t) = \frac{d}{dt}M(t) = \exp(\alpha + \beta t)$$

$$\lambda(t) = Kmt^{m-1}$$

複数の計数過程のモデル化

- 複数の計数過程において，特異な意味を持つnuisance パラメータと共通のトレンド・パラメータを使ってモデル化すると，nuisance パラメータがトレンド・パラメータの推測に関して精度を低下させてしまうことが知られている．

$$\lambda_k = \exp(\alpha_k + \beta t)$$

条件付き尤度の利用

- nuisance パラメータの影響を取り除く方法として条件付き尤度が考えられているが，条件付き尤度がnuisance パラメータを含まない強度モデルを与えるための必要十分条件は以下の通りである．(Kamakura, 1996)

$$\lambda(t;\alpha,\beta) = h(\alpha)g(\beta;t) \quad a.s.$$

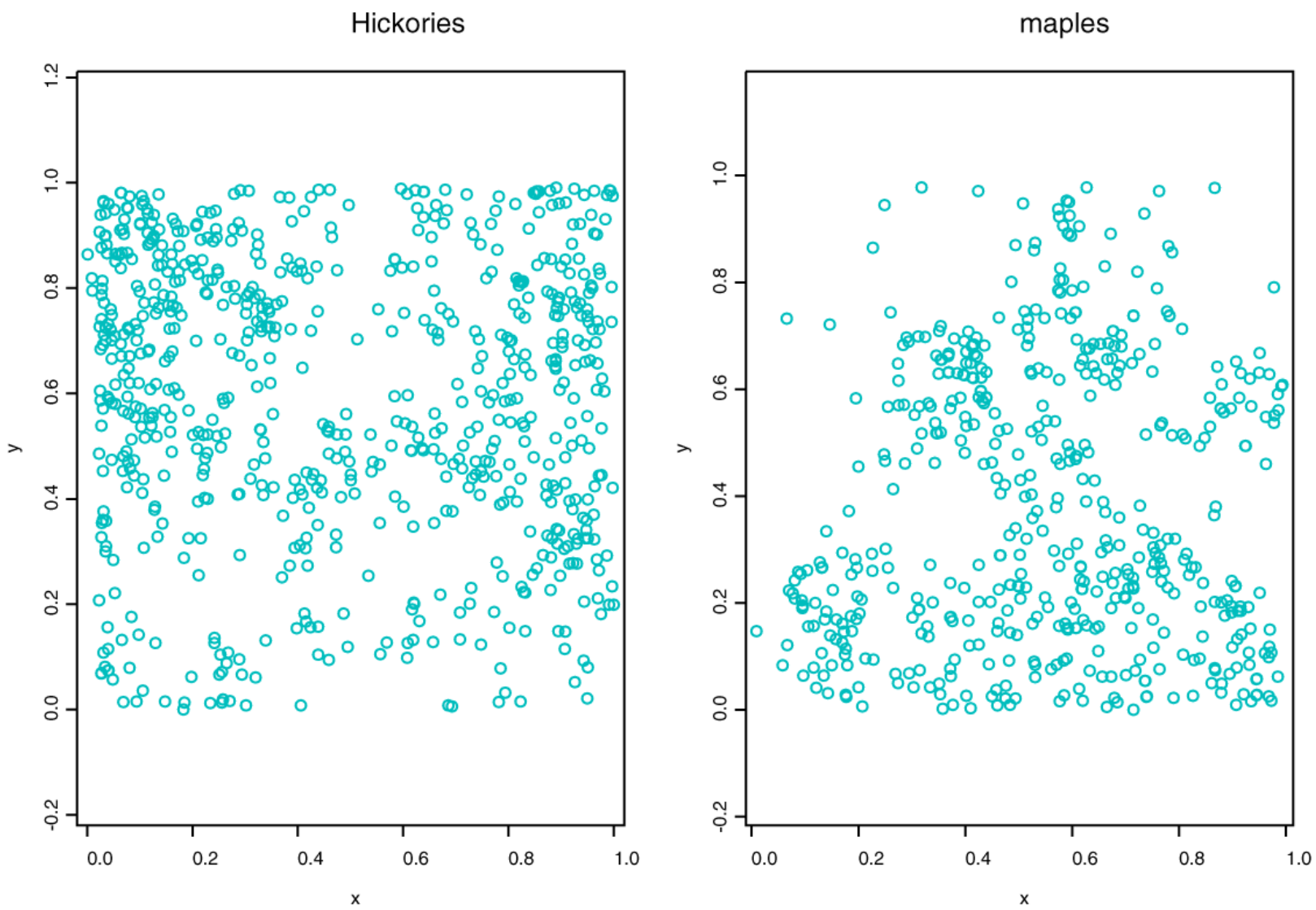
空間データとは

- 空間の要素は
  - 点，線，領域，面
- 点
  - *地震の震源，店舗位置，歩行者の位置*
- *線*
  - *活断層，河川，峡谷，稜線*
- *領域*
  - *土砂災害，洪水，津波の被害範囲*

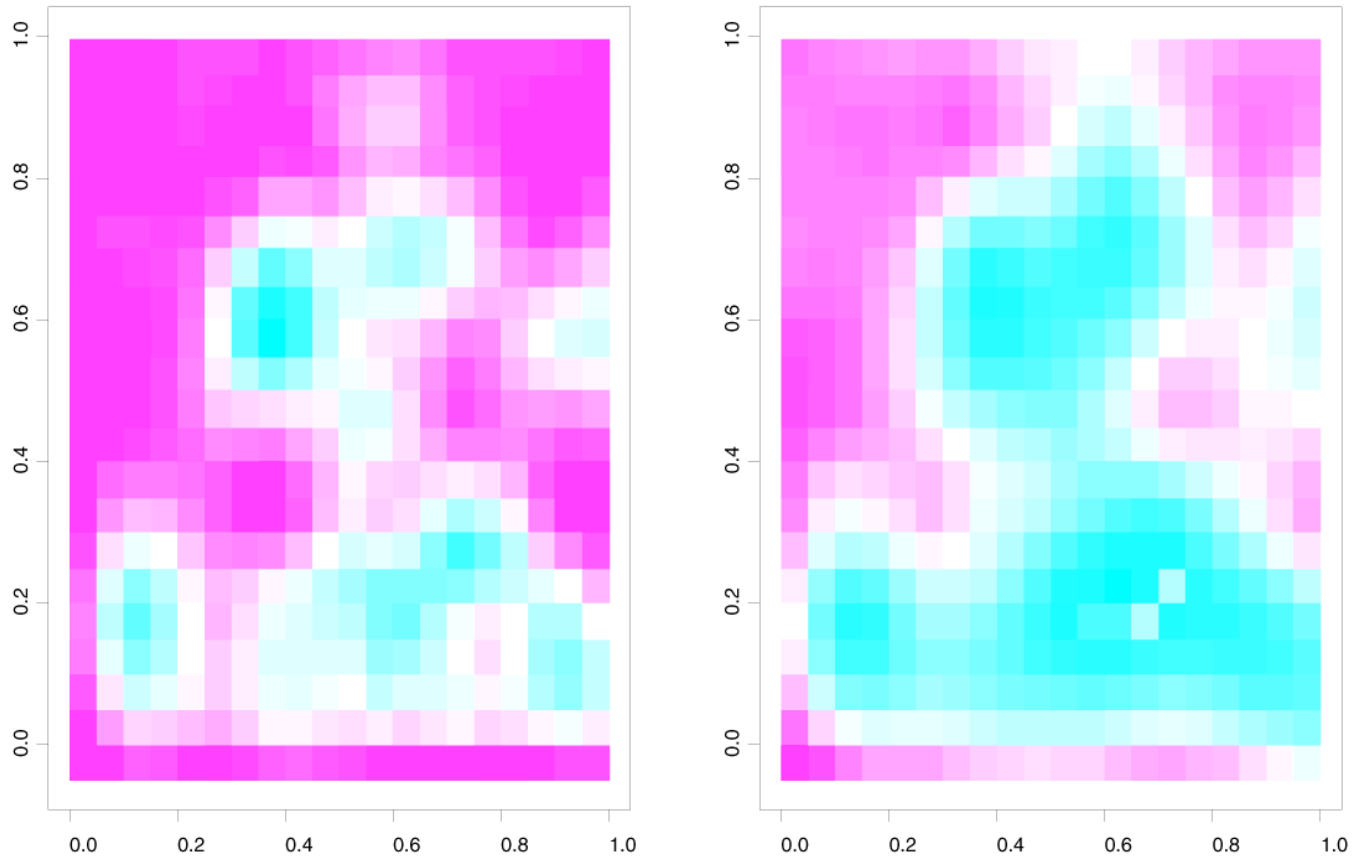
点配置の問題

- 店舗を平面上の点とみなす
- 空間パンタンの解析
  Upton and Fingleton(1985), Cressie (1993), Stoyan et al. (1995), Okabe et al. (2000)
- 物理学的な点配置、ポテンシャル関数の利用 (Ogata and Tanemura, 1981)

点の配置の例

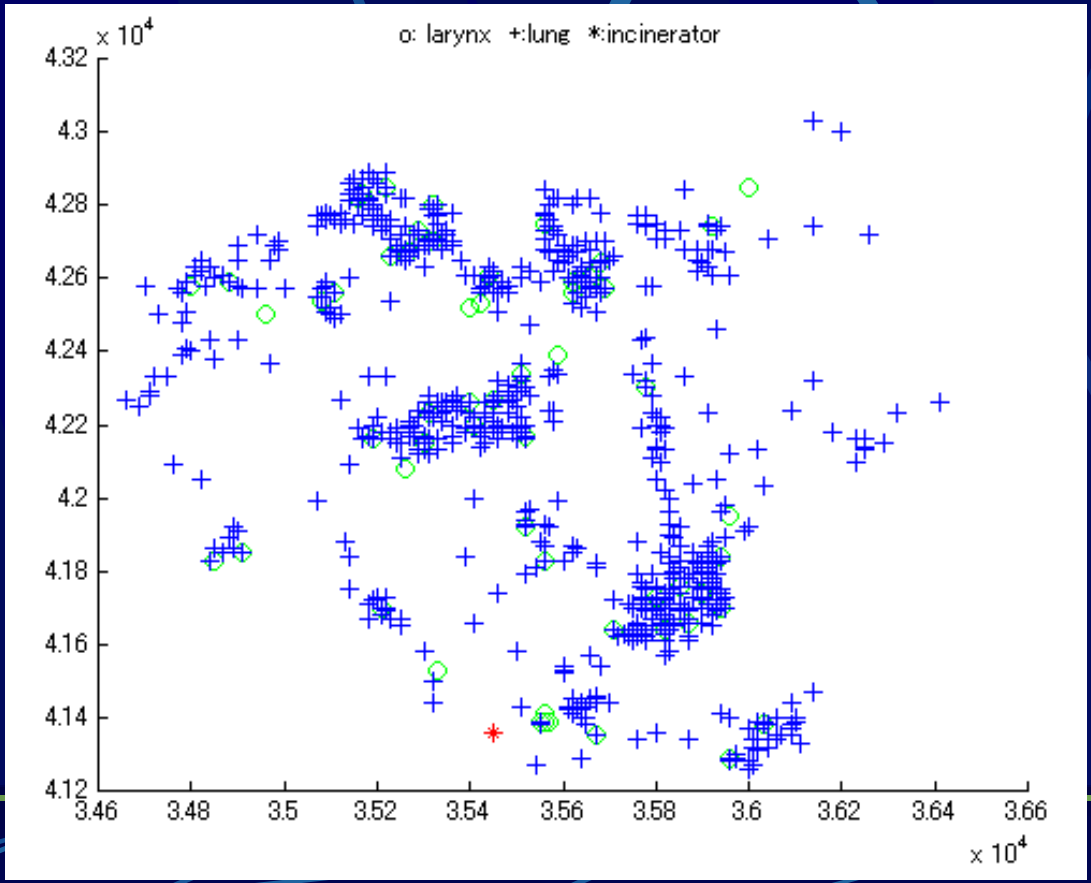


空間分布の推定

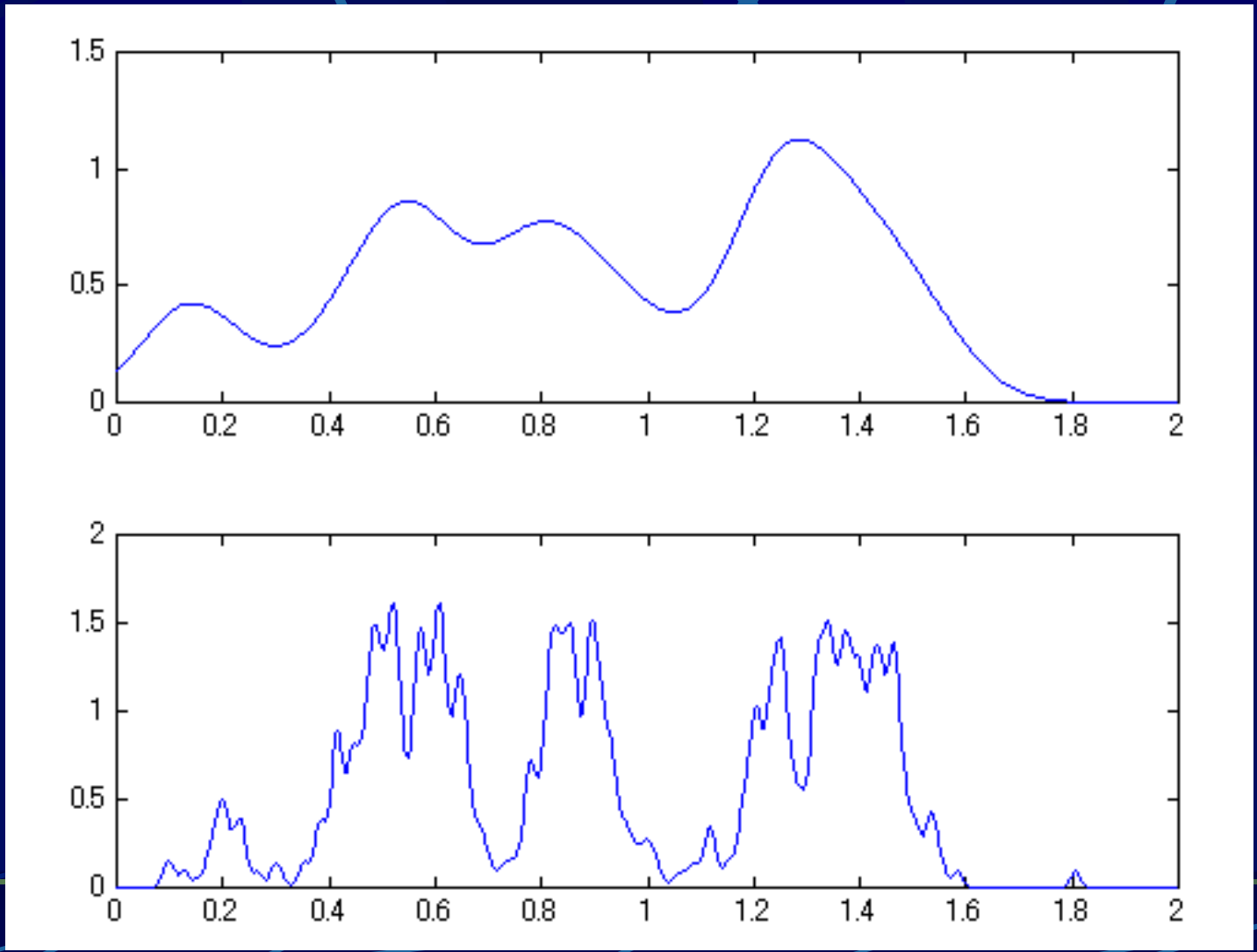


焼却場と疾病

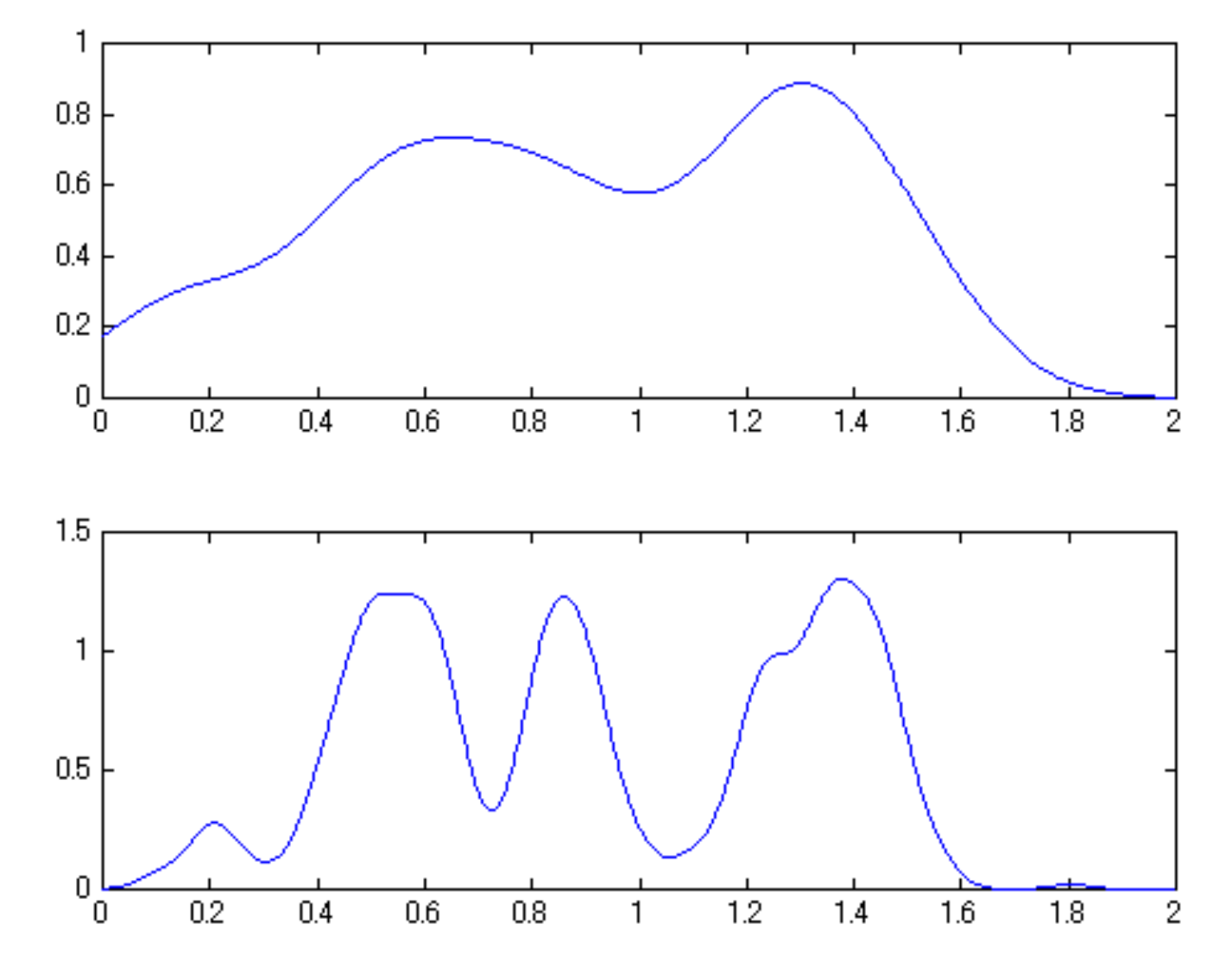
- EnglandのLancashireにおける喉頭癌と肺癌の居住者の分布と焼却場の関係



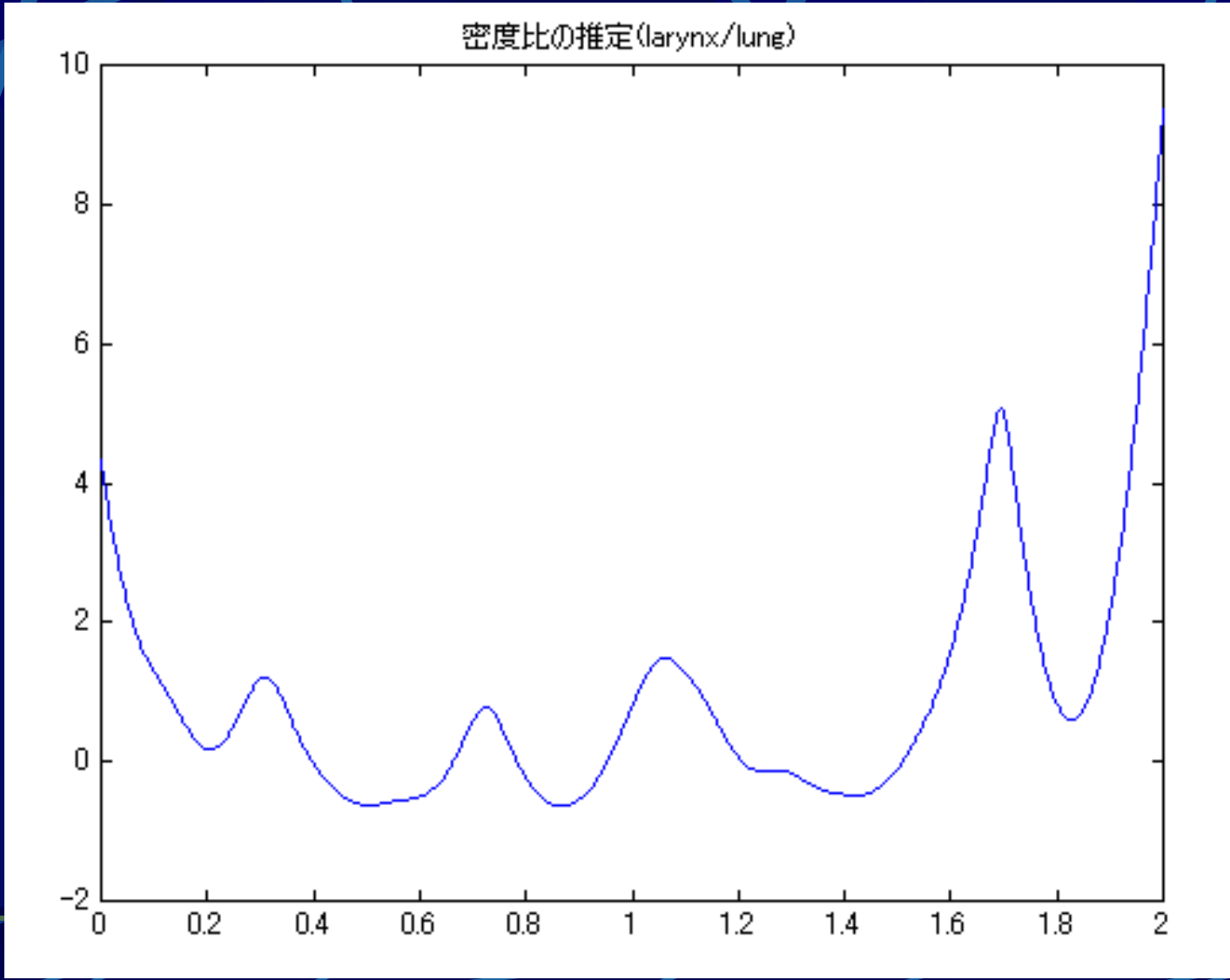
Kernel推定（1）


1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Kernel推定（2）


1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

密度比の推定



空間強度関数の推定

$$\lambda(x,y)=\sum_{k=1}^{n}\lambda_k I_k(x,y)$$

$$LL(\lambda)=-\sum_{k=1}^{K}\lambda_k|A_k|+\sum_{k=1}^{K}N(A_k)\ln\lambda_k-\ln\left(N(S)!\right)$$

$$\hat{\lambda}_k=\frac{N(A_k)}{|A_k|}$$

極値分布と災害

- 日最高気温、日最低気温のようにある標本サイズのうち最大のものあるいは最小のもののみが観測値として与えられているときに、その結果から何らかの予測を行いたい。
- 河川の堤防の建設では、洪水を起こさないようにするために、河川の水位のうち最大水位の分布が重要な役割を果たす。
- 信頼性工学では、最弱リンクモデルのように、最も弱い部分が切れるというように最小値も意味を持ってくる。

極値分布(1)

- 観測値を $X_1, X_2, \ldots, X_n$ としたとき、この値を昇順に並び替えてできる値を、$Y_1, Y_2, \ldots, Y_n$ としたとき、

$$Y_1 \le Y_2 \le \ldots \le Y_n$$

を $X_1, X_2, \ldots, X_n$ の順序統計量という。

$Y_1, Y_2, \ldots, Y_n$ の同時確率密度関数は、

$$g(y_1, y_2, \ldots, y_n) = n! f(y_1) f(y_2) \times \cdots \times f(y_n)$$

極値分布(2)

- 最大値（極値）Y_n の分布関数、確率密度関数はそれぞれ、

$$G(y_n) = \left\{F(y_n)\right\}^n$$

$$g(y_n) = nf(y_n)\left\{F(y_n)\right\}^{n-1}$$

- 最小値の分布については、

$$g(y_1) = n\left\{1 - F(y_1)\right\}^{n-1} f(y_1)$$

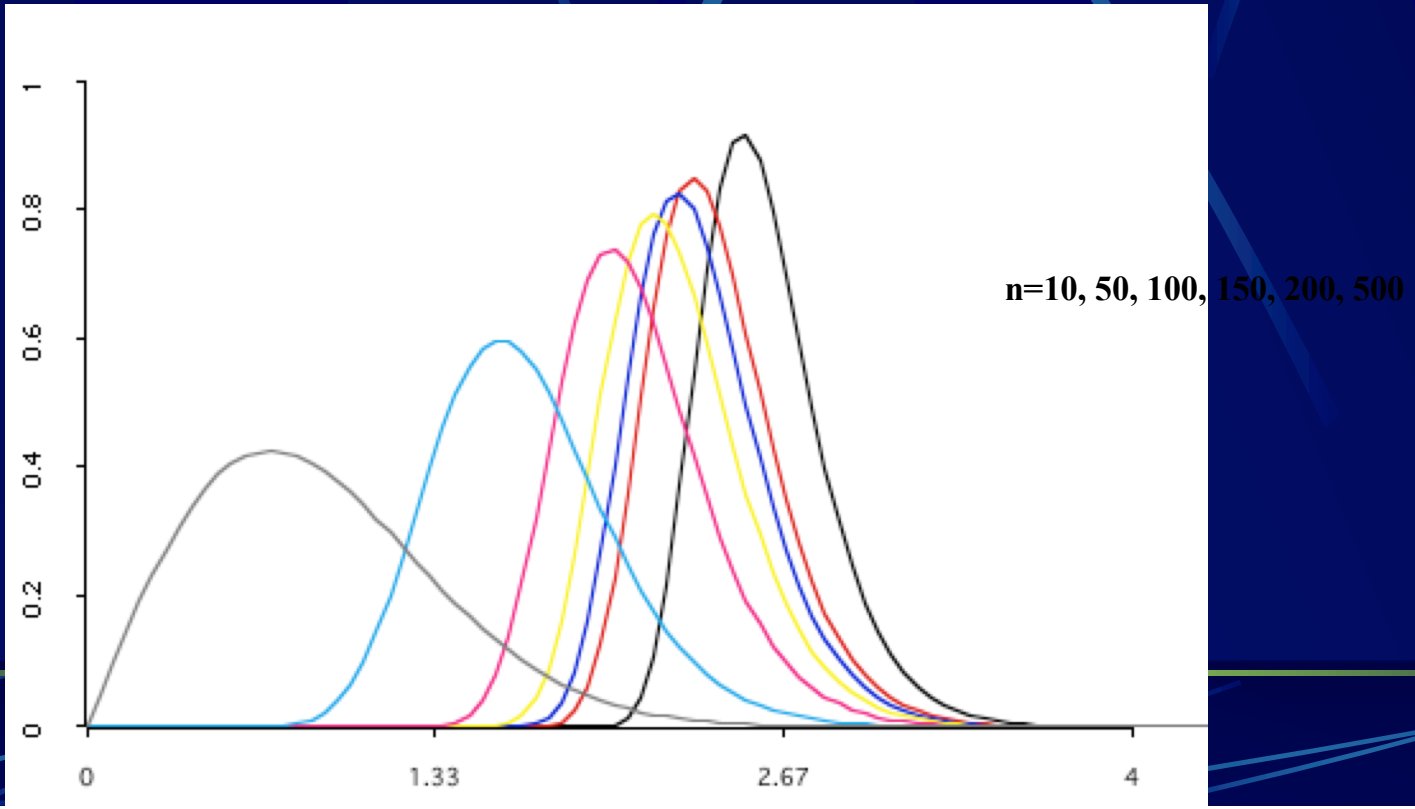
$$G(y_1) = 1 - \left\{1 - F(y_1)\right\}^n$$

極値分布(3)例

- ワイブル分布

$$f(x) = 2x\exp\left(-x^2\right)$$

$$g(y_n) = 2ny_n \exp\left(-y_n^2\right)\left\{1-\exp\left(-y_n^2\right)\right\}^{n-1}$$



Probable extreme value

- $g(y_n)$のモード: $\tilde{y}$ n個の観測値の中で最も起こりやすい極値

$$\frac{d}{dy_n} g(y_n) = 0$$

$$f'(y_n)F(y_n) + (n-1)\left\{f(y_n)\right\}^2 = 0$$

$$\frac{1}{1-F(\tilde{y}_n)} = -\frac{n-1}{F(\tilde{y}_n)} \frac{f(\tilde{y}_n)}{1-F(\tilde{y}_n)} \frac{f(\tilde{y}_n)}{f'(\tilde{y}_n)}$$

$$\frac{f(\tilde{y}_n)}{1-F(\tilde{y}_n)} = -\frac{f'(\tilde{y}_n)}{f(\tilde{y}_n)} \quad (\text{for large } \tilde{y}_n)$$

Return period (1)

$$\frac{1}{1-F(\tilde{y}_n)} \approx \frac{n-1}{F(\tilde{y}_n)} \approx n \quad (\text{for large n and } \tilde{y}_n)$$

- 上式の左辺をreturn periodという
- 洪水の予測の問題では、ある閾値uを越えるまでの平均時間（平均試行回数）と解釈できる

Return period (2)

- k回目の観測で閾値を超える確率は，p=1-F(u)とすれば，

$$P\{\tau > k\} = P\{X_1 \le u, \ldots, X_{k-1} \le u, X_k > u\}$$
$$= p(1-p)^{k-1} \ (k = 1,2,3,\ldots)$$

この期待値は，1/p=1/{1-F(u)}=T

$$\lim_{n\to\infty} P\{Y_n > \tilde{y}_n\} = \lim_{n\to\infty}\{1 - G(\tilde{y}_n)\}$$
$$= \lim_{n\to\infty}\left[1 - \{F(\tilde{y}_n)\}^n\right]$$
$$= \lim_{n\to\infty}\left[1 - \left(1 - \frac{1}{n}\right)^n\right]$$
$$= 1 - e^{-1} \approx 0.632$$

Return period (3)

- つまり，60%以上もの確率で極値が起こることになり，そのままでは実用上使用できない．十分小さな危険率αをとって，

$$G(\tilde{y}_n) = \{F(\tilde{y}_n)\}^n = 1 - \alpha$$

αが十分小さく，nが十分に大きいときには，

$$1 - F(\tilde{y}_n) = \frac{\alpha}{n}$$

極値の漸近分布

- Freshet(1927), Fisher and Tippett (1928), Gnedenko (1943)

$$\lim_{n\to\infty} P\left\{\frac{Y_n-\gamma_n}{\delta_n}\le\frac{y-a}{b}\right\}=H\left(\frac{y-a}{b}\right)$$

$$H^{(1)}\left(\frac{y-a}{b}\right)=\exp\left\{-\exp\left(-\frac{y-a}{b}\right)\right\}$$

$(-\infty < y < \infty)$

$$H^{(2)}\left(\frac{y-a}{b}\right)=\begin{cases}0 & y<a\\ \exp\left\{-\left(\frac{y-a}{b}\right)^{\alpha}\right\} & y\ge a,\quad \alpha>0\end{cases}$$

$$H^{(3)}\left(\frac{y-a}{b}\right)=\begin{cases}\exp\left\{-\left(-\frac{y-a}{b}\right)^{\alpha}\right\} & y<a,\quad \alpha>0\\ 1 & y\ge a\end{cases}$$

例

- 第１の分布に対して考えると，

$$\log\left[-\log\left\{H^{(1)}\right\}\right]=-\frac{y-a}{b}$$

極値がm個観測されたものとして，これを，
として，$y_n^{(1)}, y_n^{(2)}, \ldots, y_n^{(m)}$

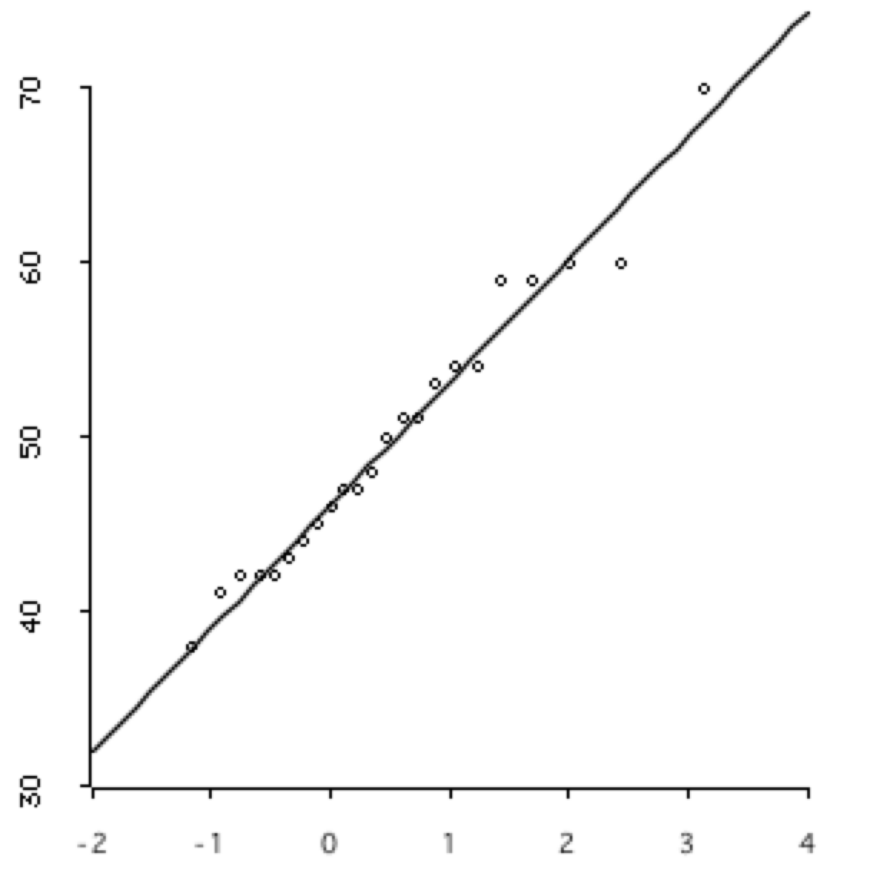
$$\begin{aligned} y &= a+bx\left[-\log\left\{-\log\left(\hat{H}^{(1)}\right)\right\}\right] \\ &= a+bX \end{aligned}$$

データの関数

(X,y)の値をm個プロットして直線をあてはめれば、
そのy切片と傾きがaとbの推定値となる

例

- New London 1873-1895年の年間最大風速(単位:mph) Changery (1982)



Return periodと風速

```
> (print-matrix (predict '(2 5 10 20 50 100 200 500) a b))
#2a(
   (      2     0.500000      48.6845   )
   (      5     0.800000      56.6853   )
   (     10     0.900000      61.9825   )
   (     20     0.950000      67.0638   )
   (     50     0.980000      73.6409   )
   (    100     0.990000      78.5695   )
   (    200     0.995000      83.4802   )
   (    500     0.998000      89.9588   )
  )
```

自然災害

- 地球温暖化
  大きな台風
  豪雨，洪水，土砂崩れ
  ひでり，砂漠化
- 地震，津波

Davison(2014) Statistics of Complex

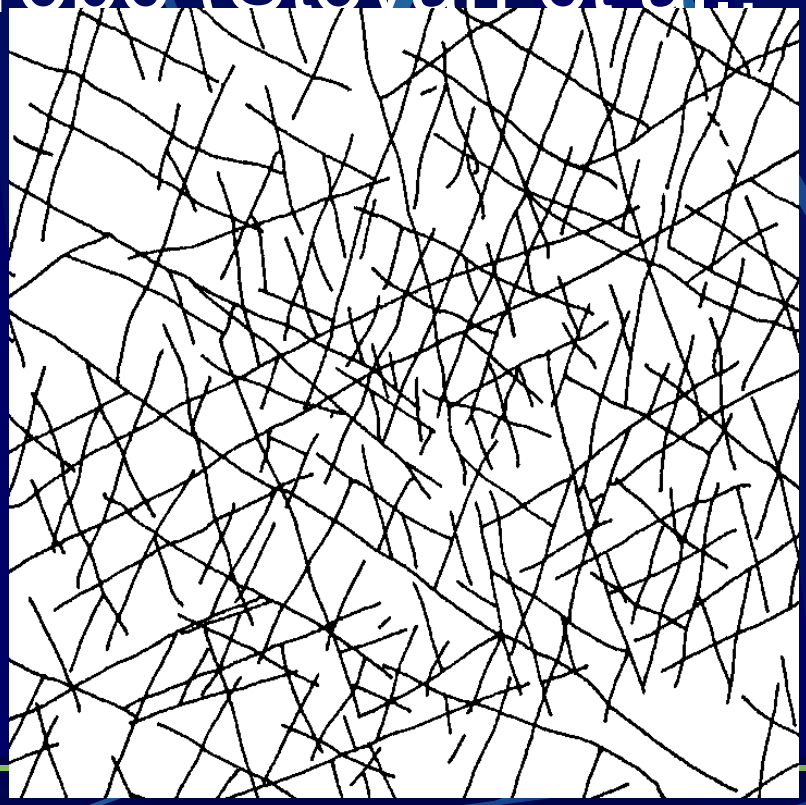
2013.1.1-20	Winter Weather	インド，バングラ	329
2013.1-2	洪水	南アフリカ	175
2013.4	地震	中国	196
2013.6	洪水	インド，ネパール	6748
2013.7	洪水	インド	174
2013.8-9	洪水	パキスタン	234
2013.9.24-28	地震	パキスタン	825
2013.10.15	地震	フィリピン	222
2013.11.7-10	台風30号 Haiy	フィリピン，ベト	8,000
2013.11.9	TS Three	ソマリア	440

空間上のイベントデータ

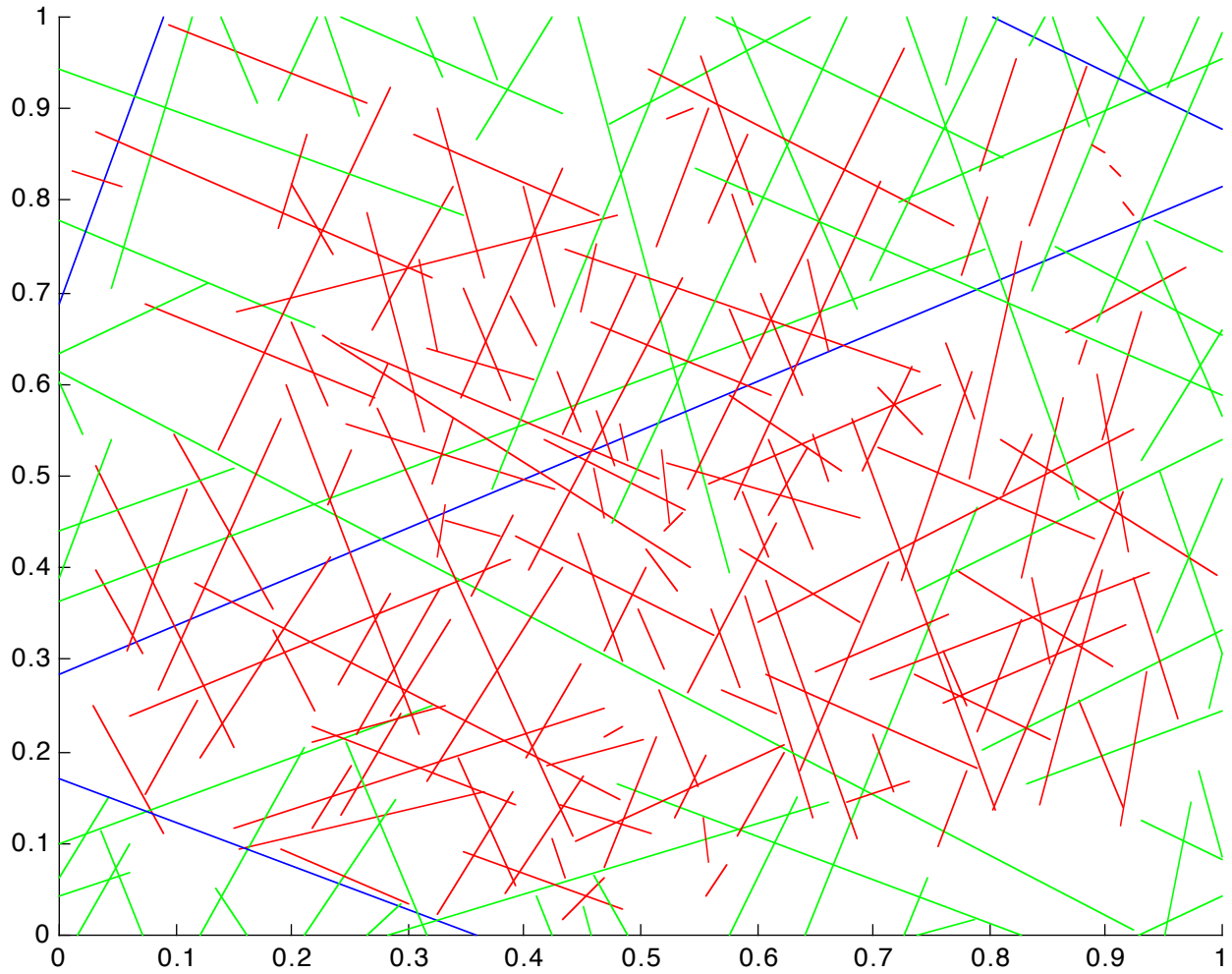
- 地震
- 活断層
  線分の空間分布（点から線へ）
- 地滑り，土砂崩れ
  点から面へ
- 空間上のイベント位置の推定
  テロ

活断層のデータ(1)

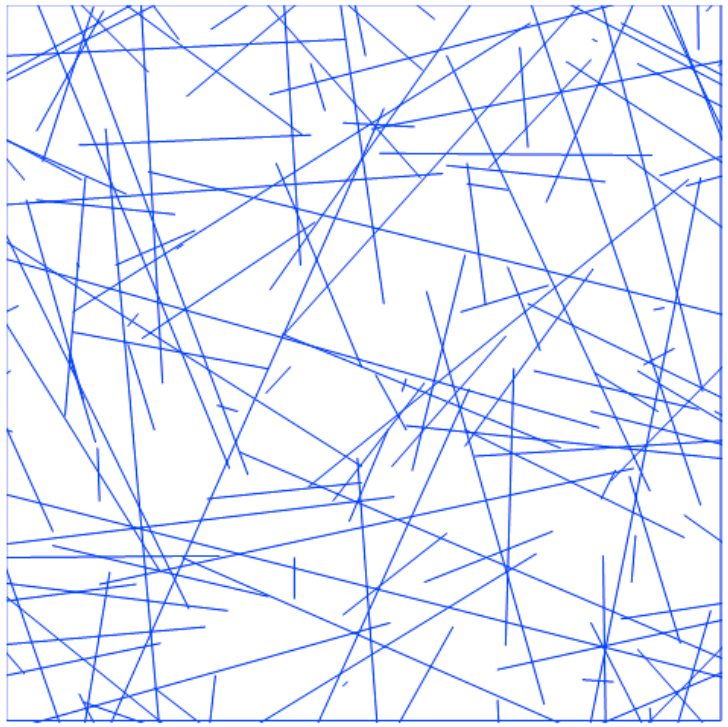
- Weber(1977):Linear fault zones in central Europe (Stovan et al., 1987)



活断層のデータ(2)



ランダムな線分



mu	TRUE	all	complete
0.05	10.051	10.622	9.417
0.10	9.923	11.142	8.688
0.15	10.035	11.773	8.250
0.20	9.829	12.185	7.634
0.25	9.861	12.695	7.169
0.30	9.959	13.209	6.754
0.35	10.087	13.691	6.406
0.40	10.018	13.953	6.069
0.45	9.988	14.351	5.750
0.50	9.999	14.479	5.464

共変量を持つ強度モデル

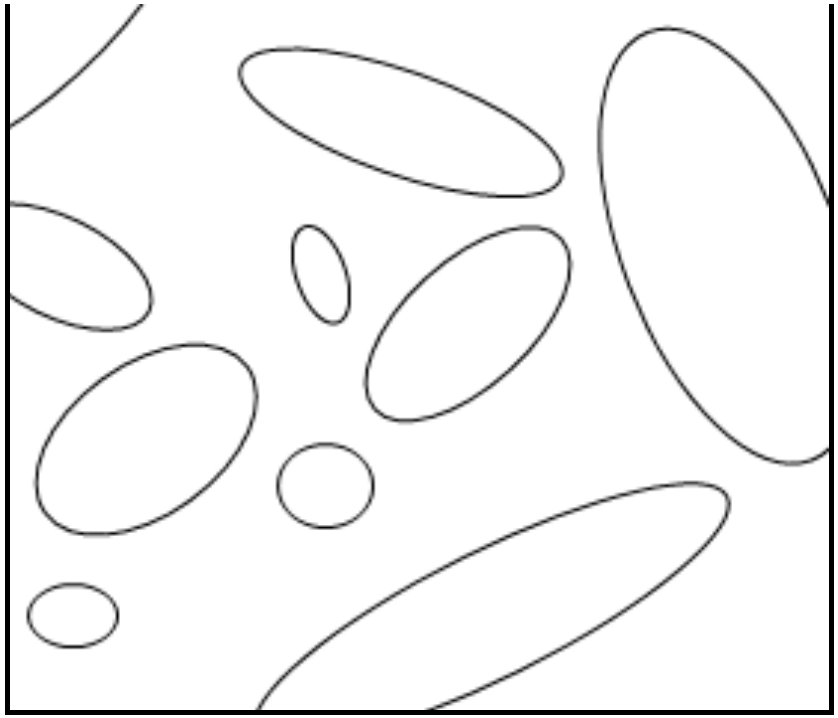
- 店舗(xi, yi) (I=I, …, n)のボロノイ領域で強度関数が次式のように一定な共変量を持つ強度モデルを考える.

$$\lambda(x_i, y_i; \mathbf{z}_i) = \exp(\mathbf{z}_i \beta)$$

$$\mathbf{z}_i = (1, z_{i1}, z_{i2}, \ldots, z_{ip})$$

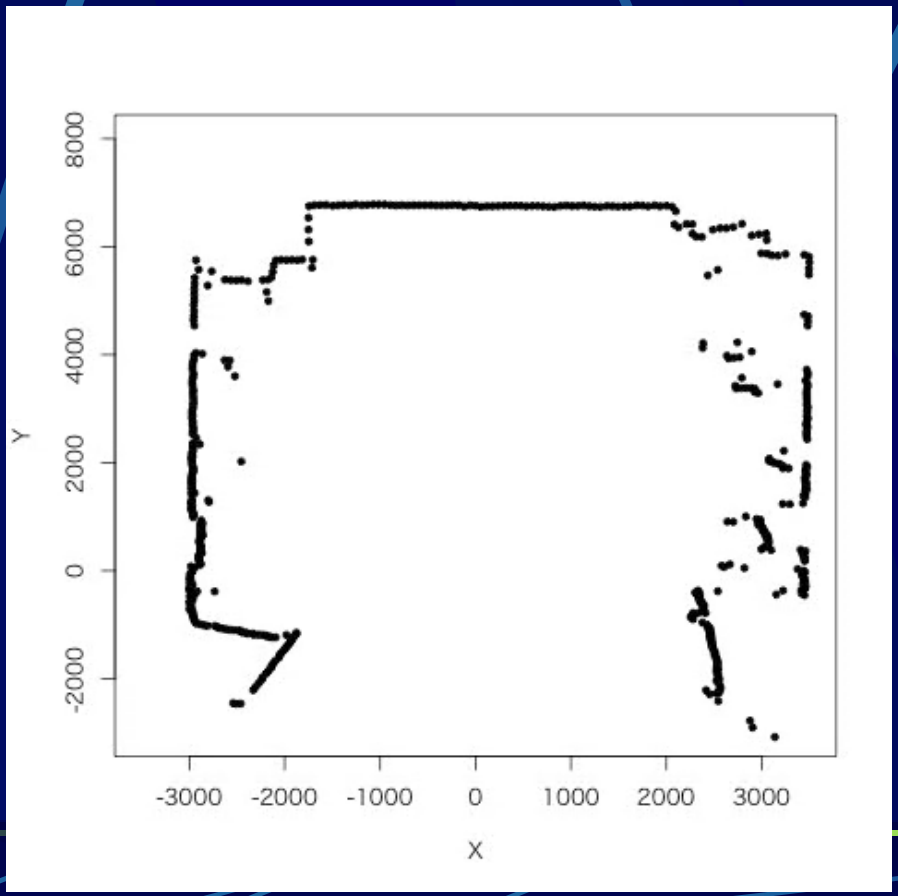
ランダムな楕円

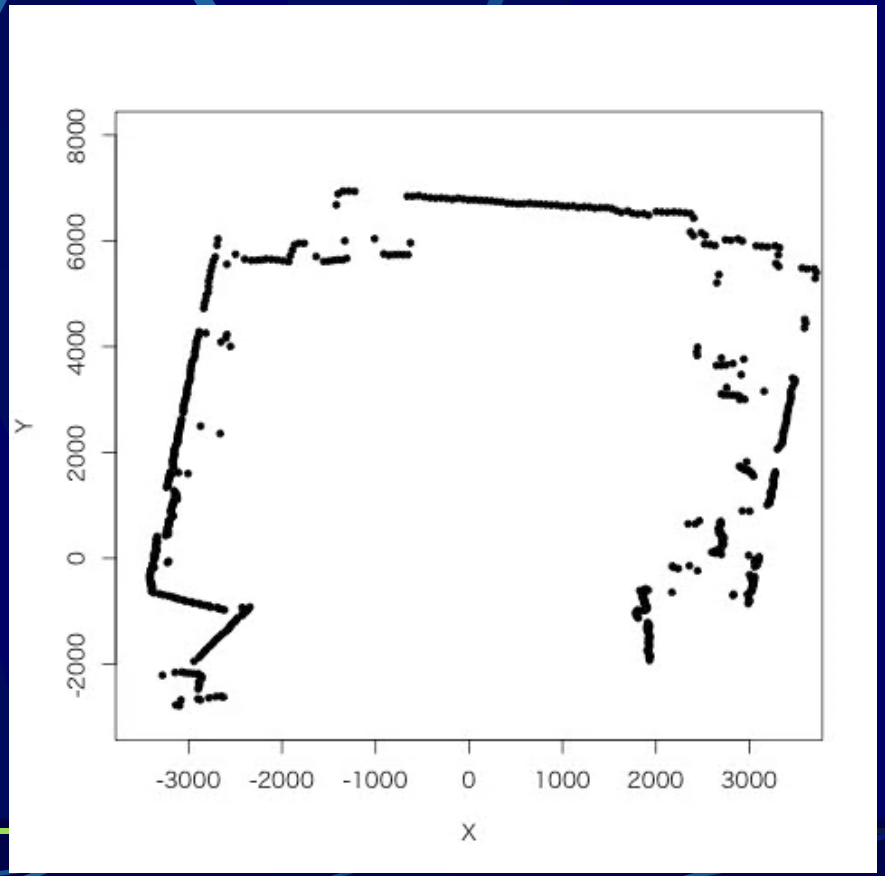
- 点から面へ



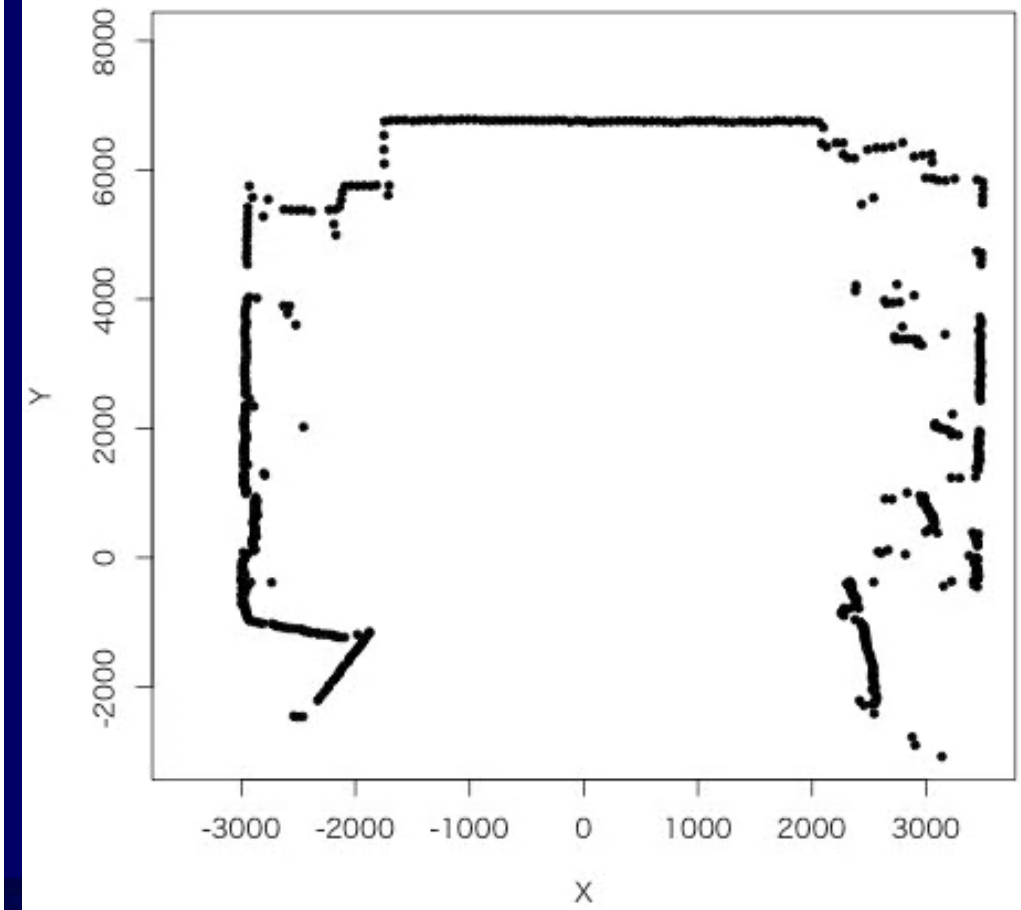
位置データ

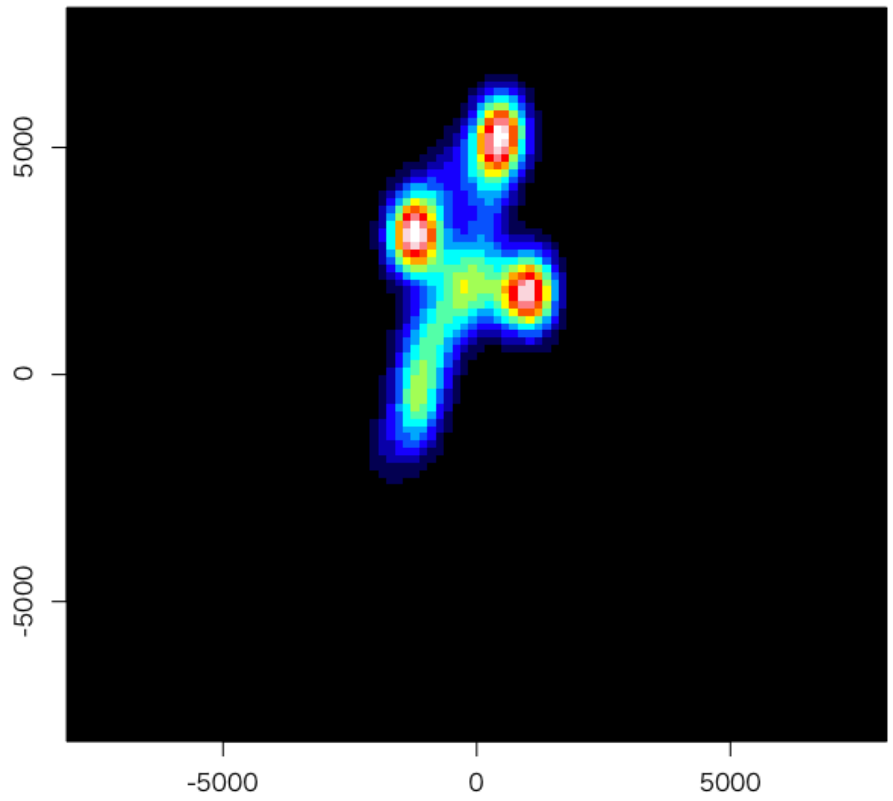
- 人の行動パターンの認識と位置データ





歩行の位置データ


8000
6000
4000
2000
0
-2000
Y
-3000
-2000
-1000
0
1000
2000
3000
X


5000
0
-5000
-5000
0
5000

災害予測精度の統計的評価法

- Mr. サンデー（フジテレビ）の取材
  - 村井先生の地震予測はあたっているといえるか.
  - データはあるのか.
  - メルマガのデータ

予測データ

東北地方および関東地方の太平洋岸は要注意（1/8）

東北および関東の太平洋岸は要注意 （1/15）

東北および関東の太平洋岸は要注意（1/22）

東北地方・関東地方の太平洋岸は要注意（1/29）

（2/5）東北地方は要注意

（2/12）東北、関東の太平洋岸は要注意

（2/19）東北・関東の太平洋岸は引き続き要注意「震度4～5程度」

（2/26）東北・関東の太平洋岸は引き続き要注意

（3/5）東北・関東の太平洋岸は引き続き要注意

（3/12）東北・関東の太平洋岸は引き続き要注意

（3/19）東北・関東の太平洋岸は引き続き要注意

2014/1/1

災害予測の精度

- データとしては何が必要か
  - 複数回の予測結果
  - 結果の評価値
- 予知，予測，予報
  - 確率予報
  - 空間的，時間的確率予報をどのように扱うか.
  - 検定論的評価

予測結果データ

	期間	地域	規模	AND	OR
1	1	0	0	0	1
2	1	1	1	1	1
3	1	1	0	0	1
4	1	1	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮
96	1	1	0	0	1
# 1	49	38	12	9	50
# 0	34	45	70	73	32
# NA	13	13	14	14	14

ポアソン過程

$$P\{N(T)=n\}=\frac{\Lambda(T)^n}{n!}\exp\{-\Lambda(T)\}$$

$$\Lambda(T)=\int_0^T \lambda(u)\,du$$

$\lambda(t)=\lambda_0$ とすると

$$P\{N(T)=n\}=\frac{(\lambda_0 T)^n}{n!}\exp(-\lambda_0 T)$$

ノンパラメトリックな最尤推定量としては，$\hat{\lambda}_0=\frac{1}{T}$

統計的仮説検定

$\lambda(t) = \lambda_0$ として

$$\begin{aligned} P\{N(T)=0\} &= P\{N(T)=1\} = \exp(-1) \\ P\{N(T)>0\} &= 1-\exp(-1) \end{aligned}$$

となる。よって、$p_1 = P\{N(T)=1\}, p_2 = P\{N(T)>0\}$ として、帰無仮説

$$H_0 : p = p_1$$
$$H_A : p = p_2$$

を 検定する つまり、

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

が p_1 を含むかどうかを計算する。$\hat{p}$ はデータから計算される値。$p_1 = \exp(-1)$。

結果

$$p_1 = e^{-1} = 0.367879, p_2 = 1 - e^{-1} = 0.632121$$

3 カラム独立

$$\hat{p} = \frac{r}{n} = \frac{49 + 38 + 12}{96 \times 3} = \frac{99}{288} = 0.34375, \ (0.2888, 0.3986)$$

AND

$$\hat{p} = \frac{r}{n} = \frac{9}{96} = 0.09375, \ (0.060, 0.127)$$

OR

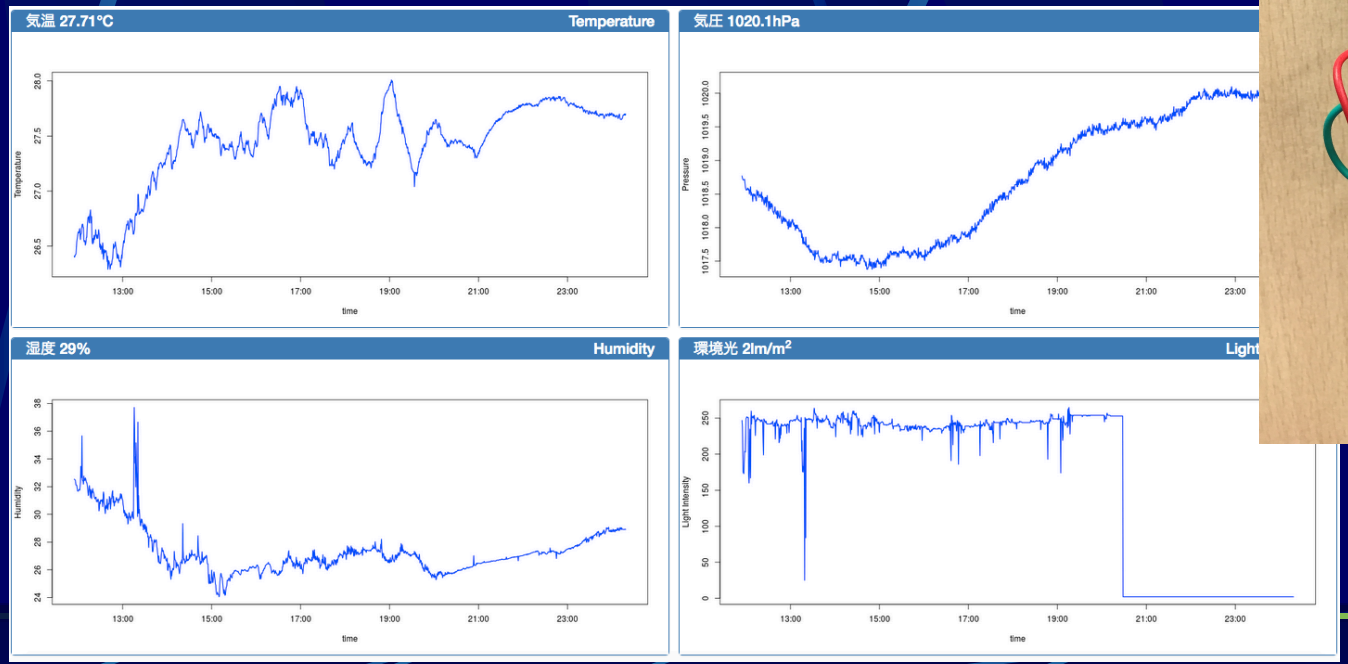
$$\hat{p} = \frac{r}{n} = \frac{50}{96} = 0.520833, \ (0.463, 0.579)$$

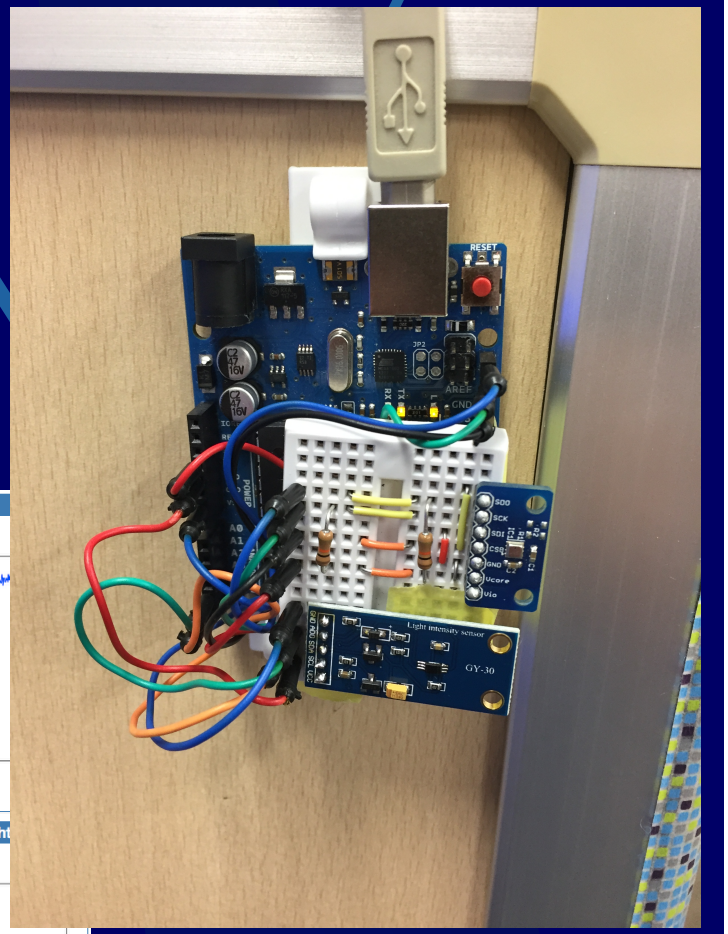
OR+NA

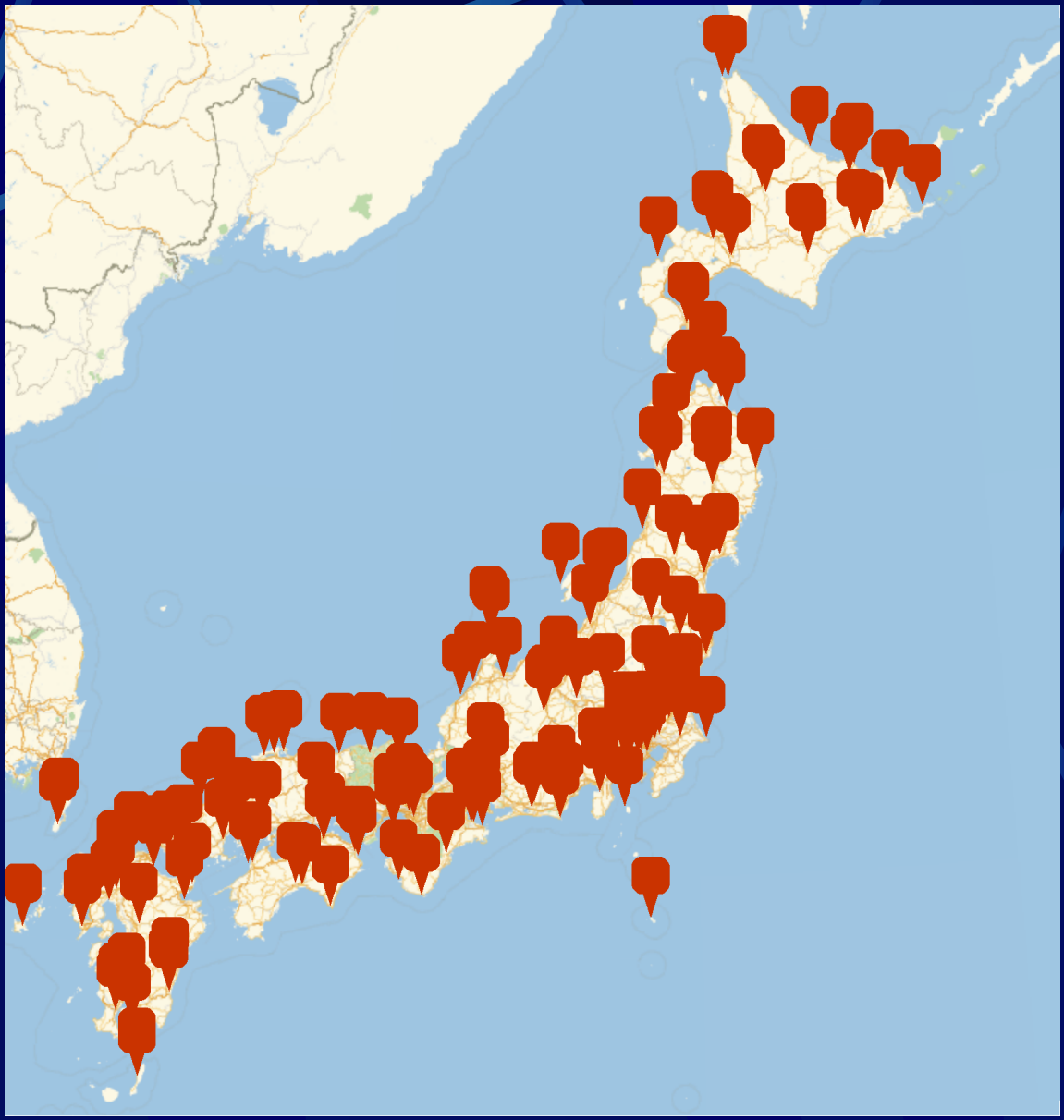
$$\hat{p} = \frac{r}{n} = \frac{64}{96} = 0.66667, \ (0.612, 0.721)$$

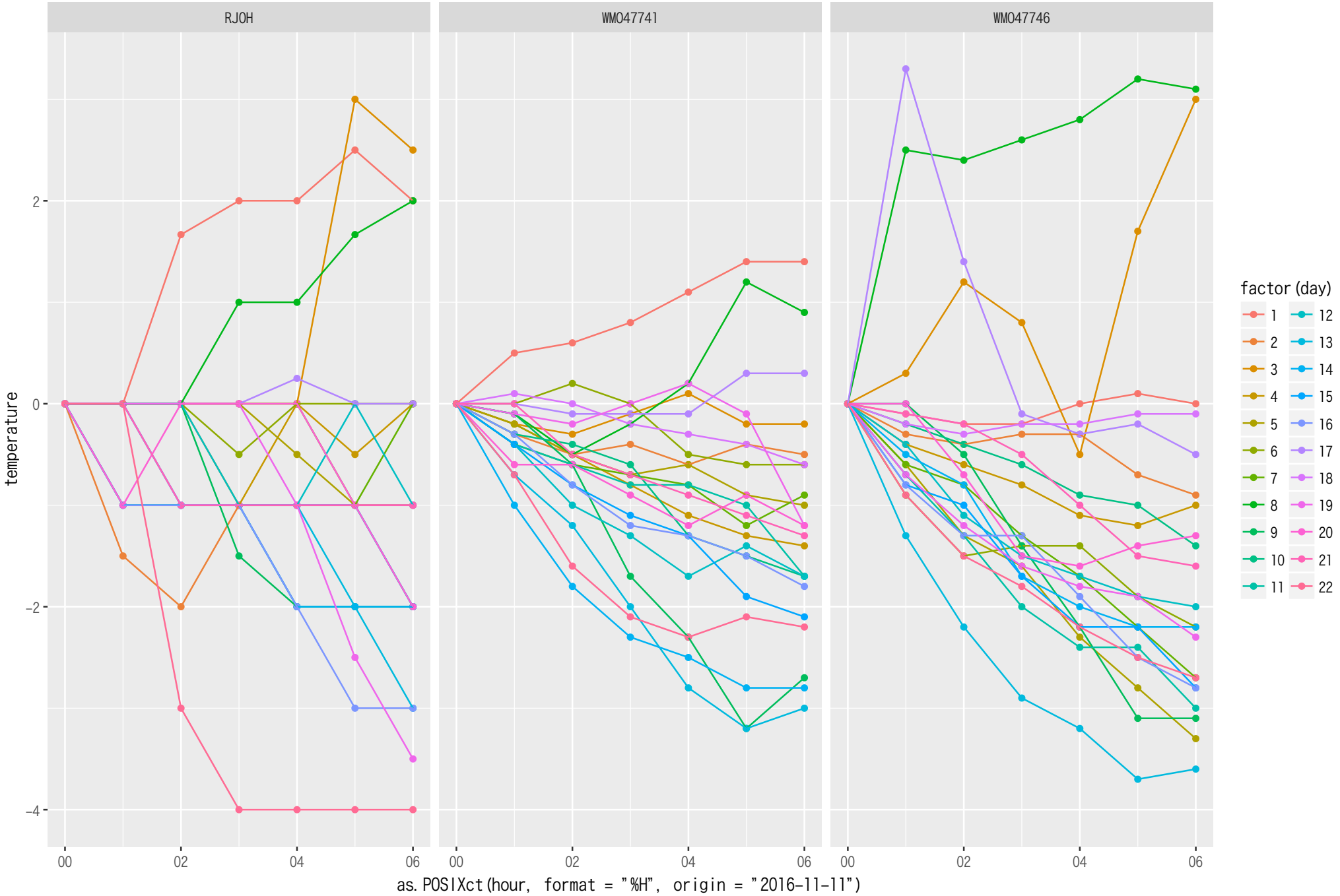
IoTの活用

- 気温の変化と地震

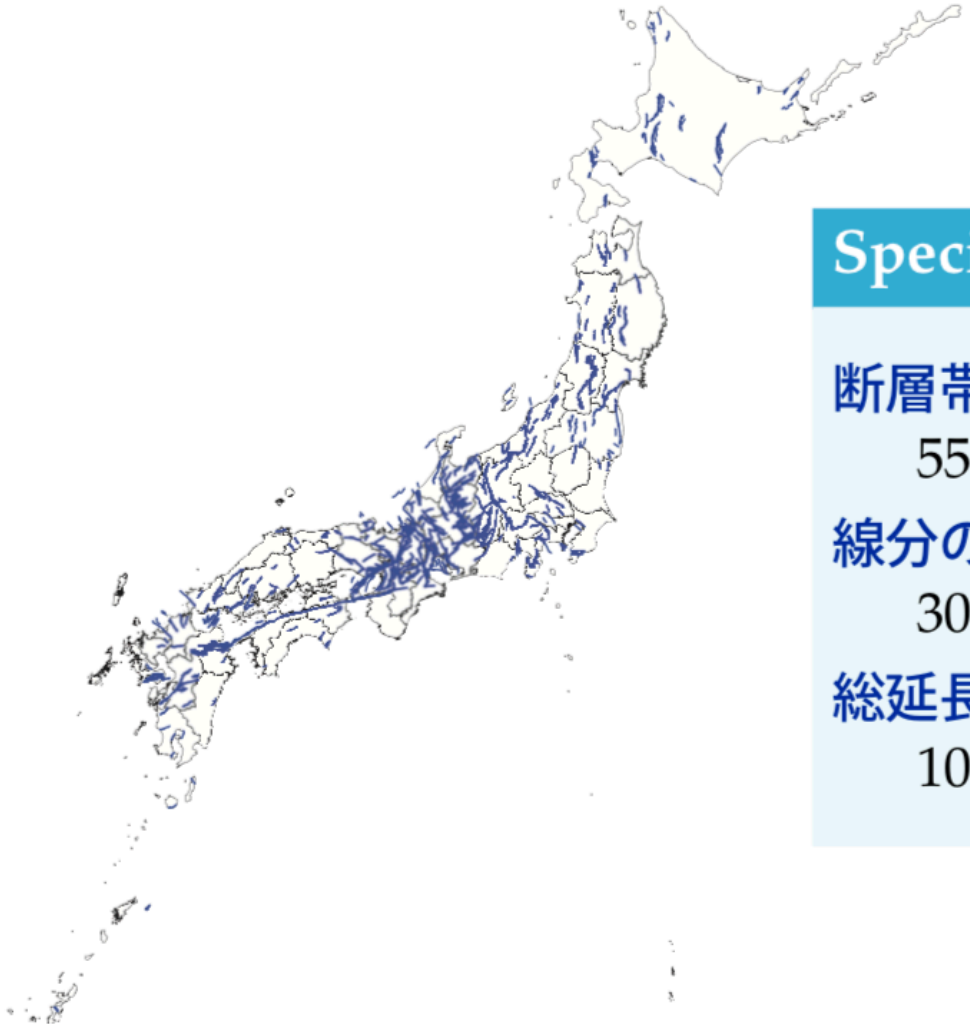







RJOH
WMO47741
WMO47746
temperature
2
0
−2
−4
00
02
04
06
as.POSIXct(hour, format = "%H", origin = "2016-11-11")
factor (day)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

日本の活断層



Specifications

断層帯の数

559

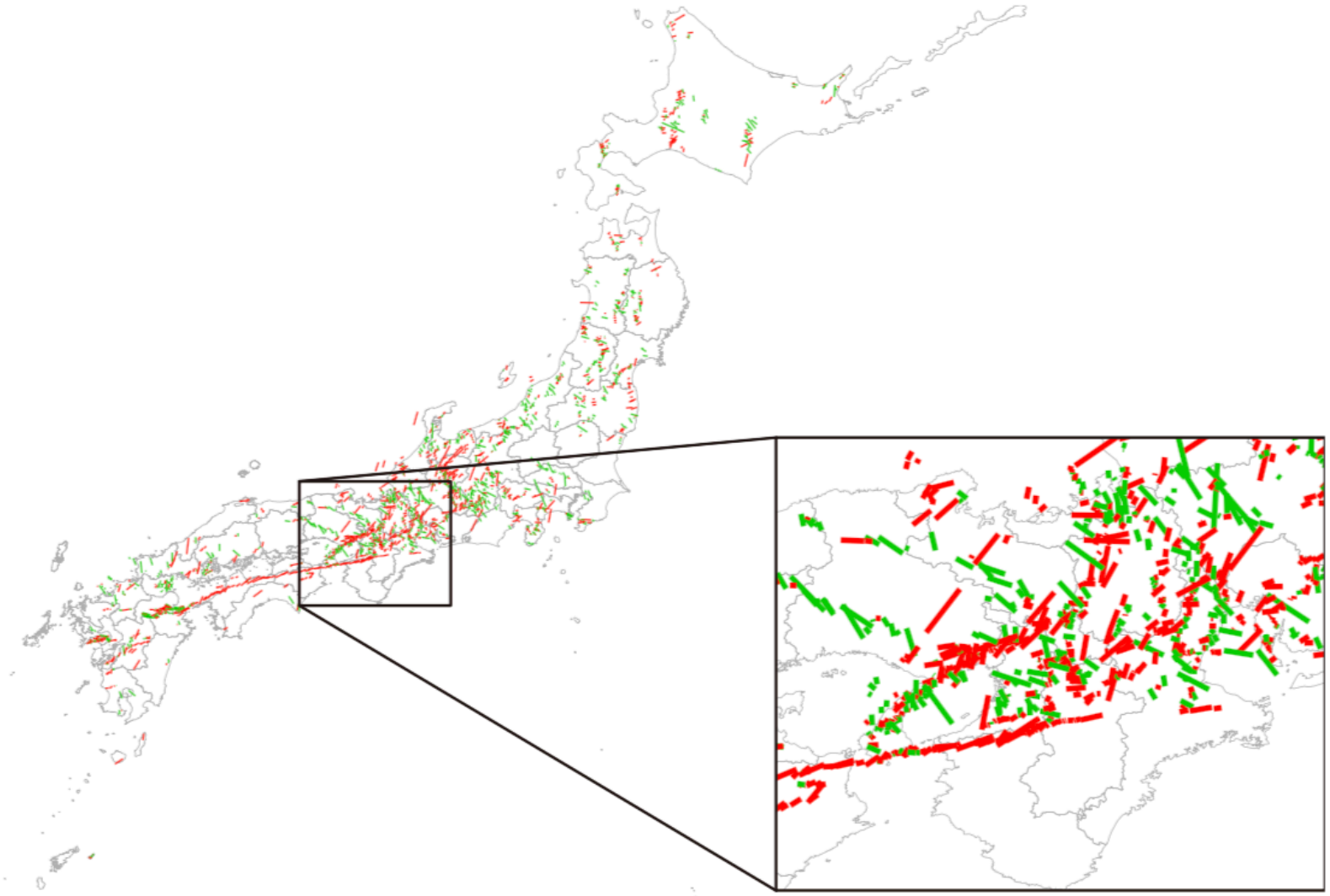
線分の要素数

3069

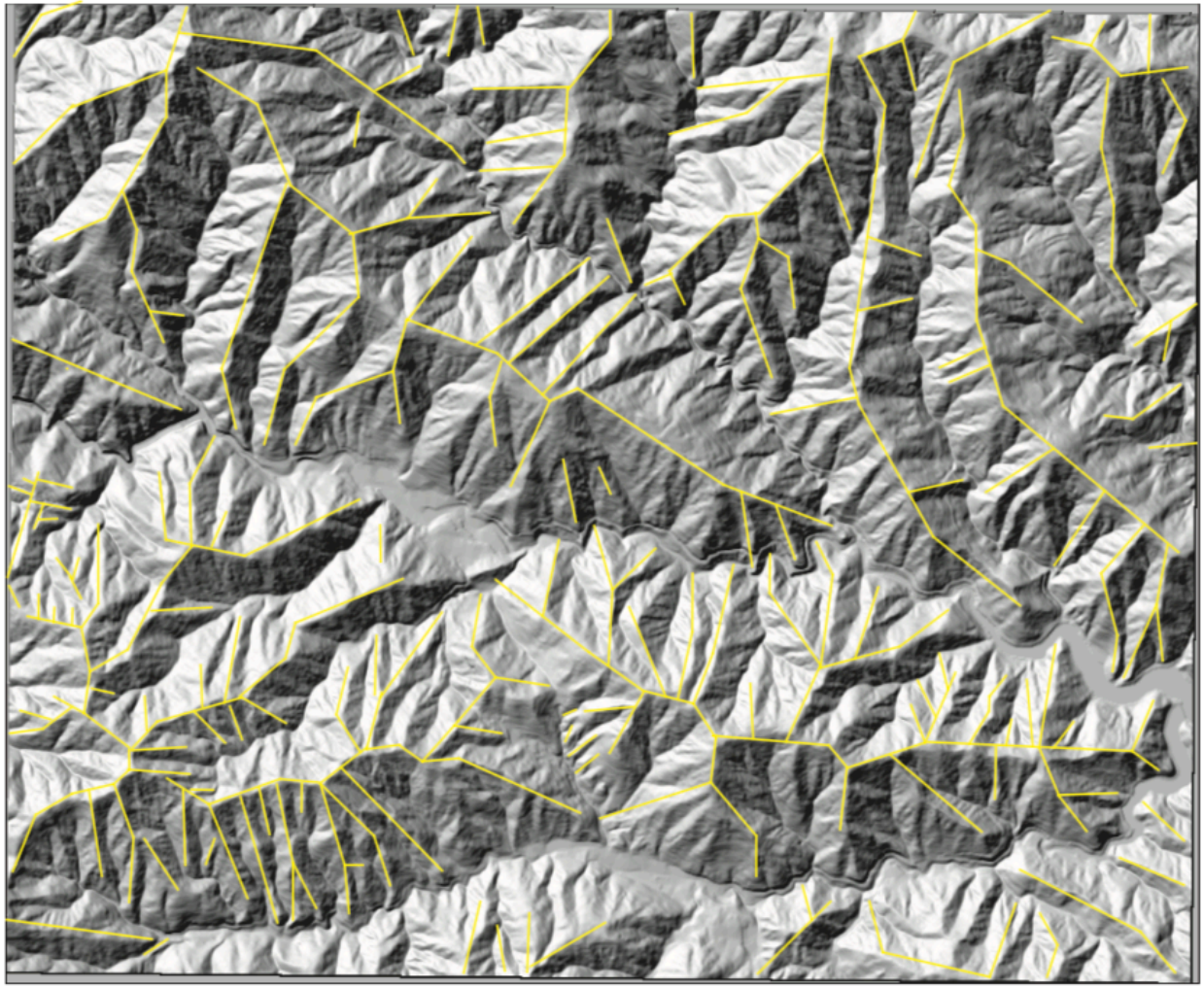
総延長

10803[km]

p 値:0.0007542133 0.0001545573



多摩地方の山々の尾根線

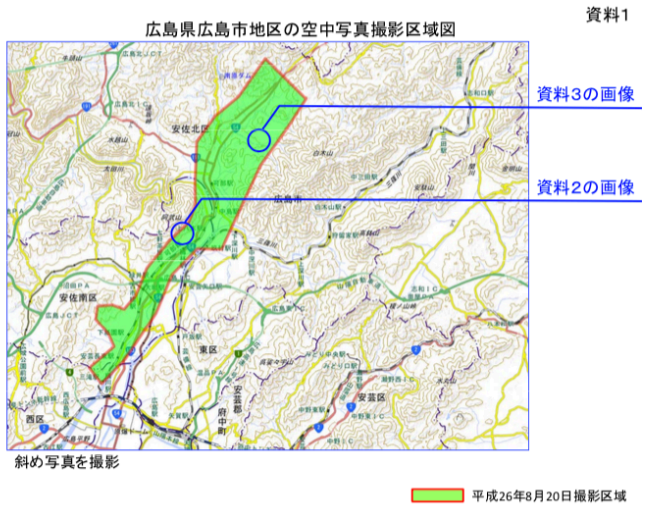


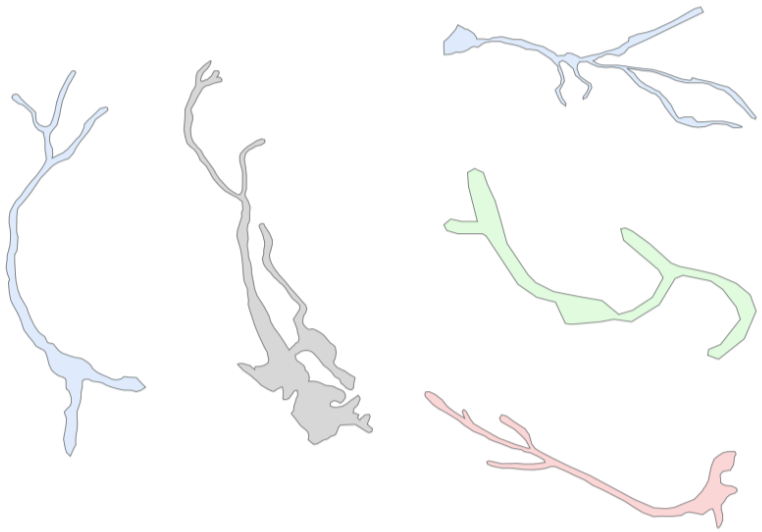
土砂崩れ

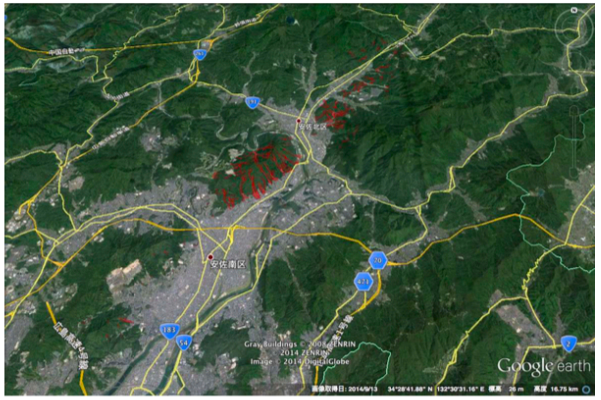
- 平成 26 年 (2014 年)8 月豪雨による広島県土砂被害
- 死者 74 人・負傷者 44 人
- うち、死者 52 人・負傷者 20 人は八木地区に集中
- （旧名は、八木蛇落地悪谷（やぎじゃらくじあしだに））
- 防災の観点から、被害状況の把握は重要

国土地理院のデータ

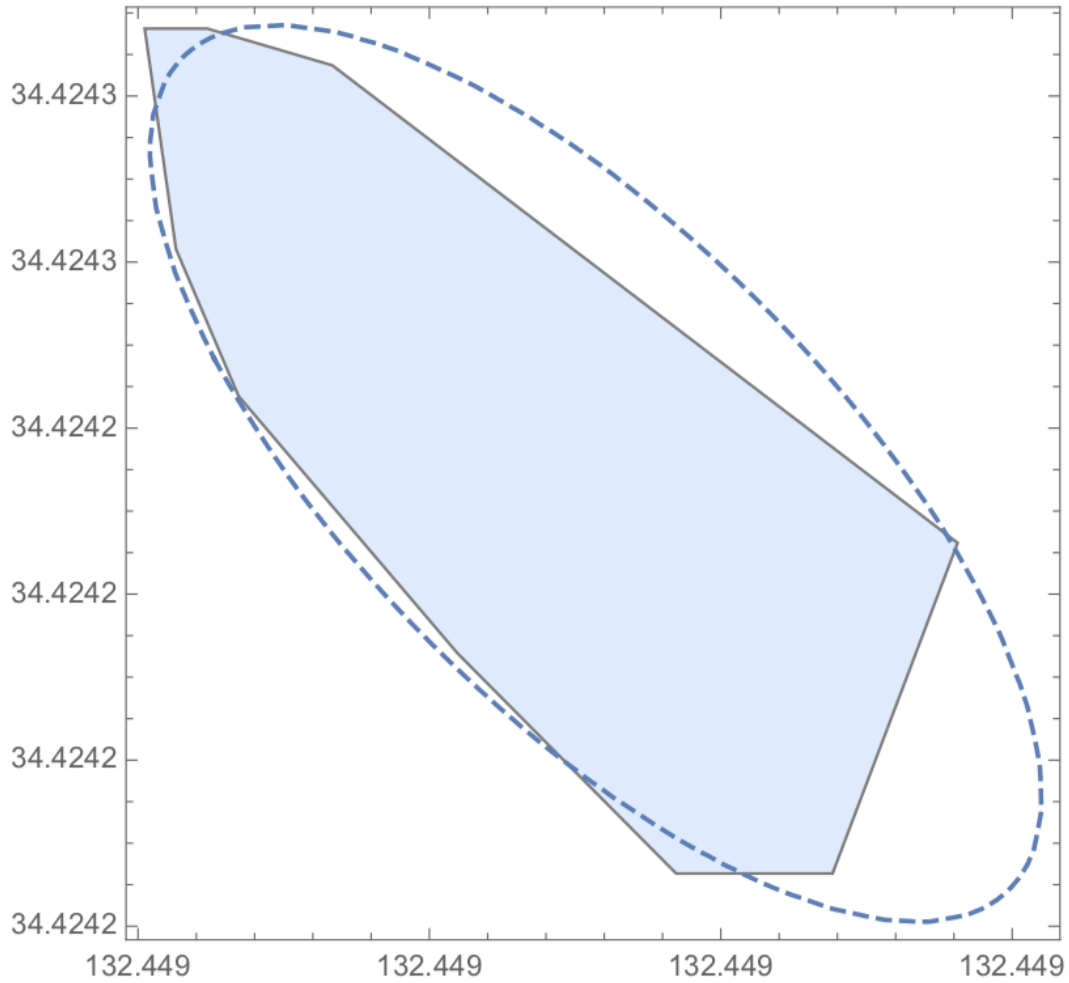
- 国土地理院による空中写真
- KML ファイルの提供
- 緯度経度の座標によるポリゴンデータ





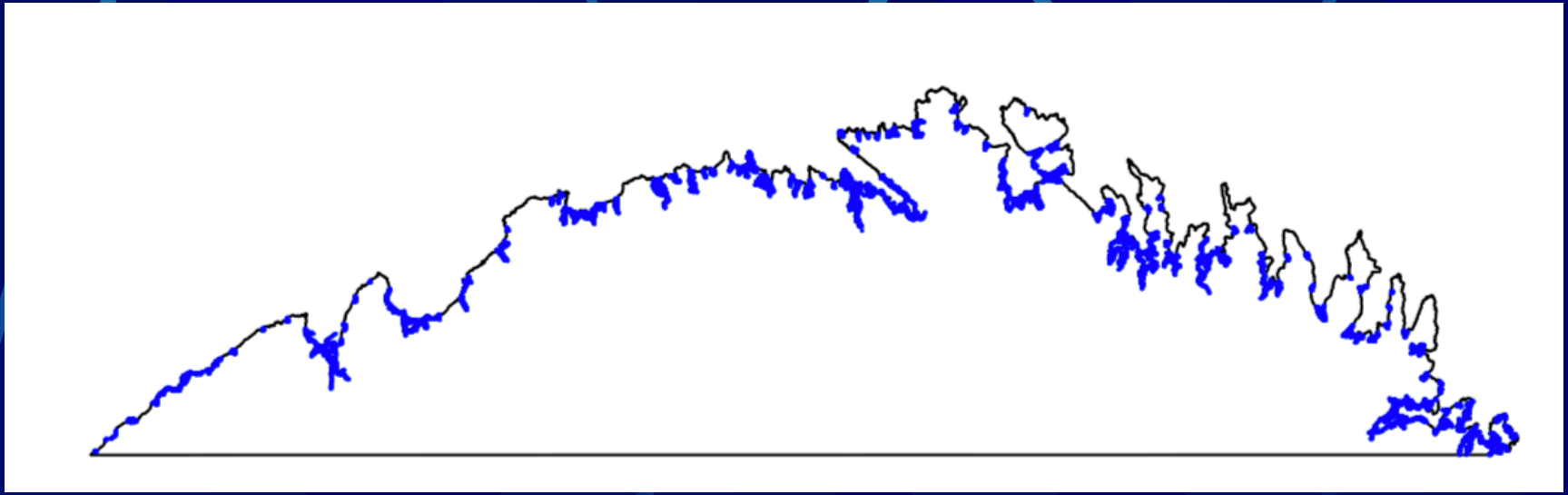


楕円による領域近似とサマライズ



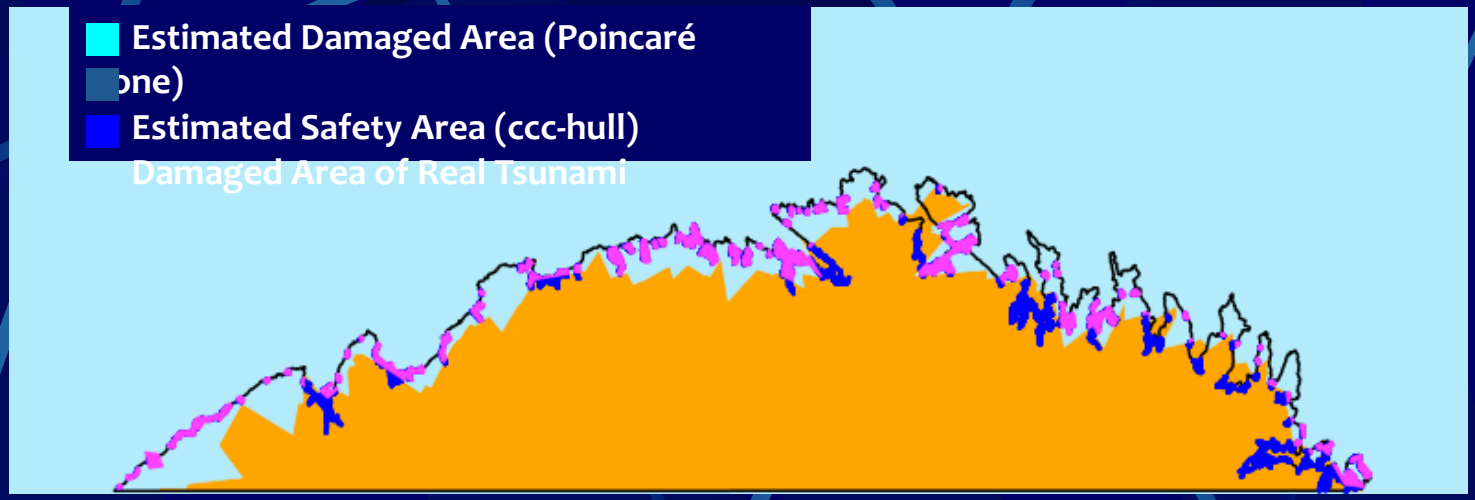
津波遡上の評価

岩手県地域と実際の津波被害範囲

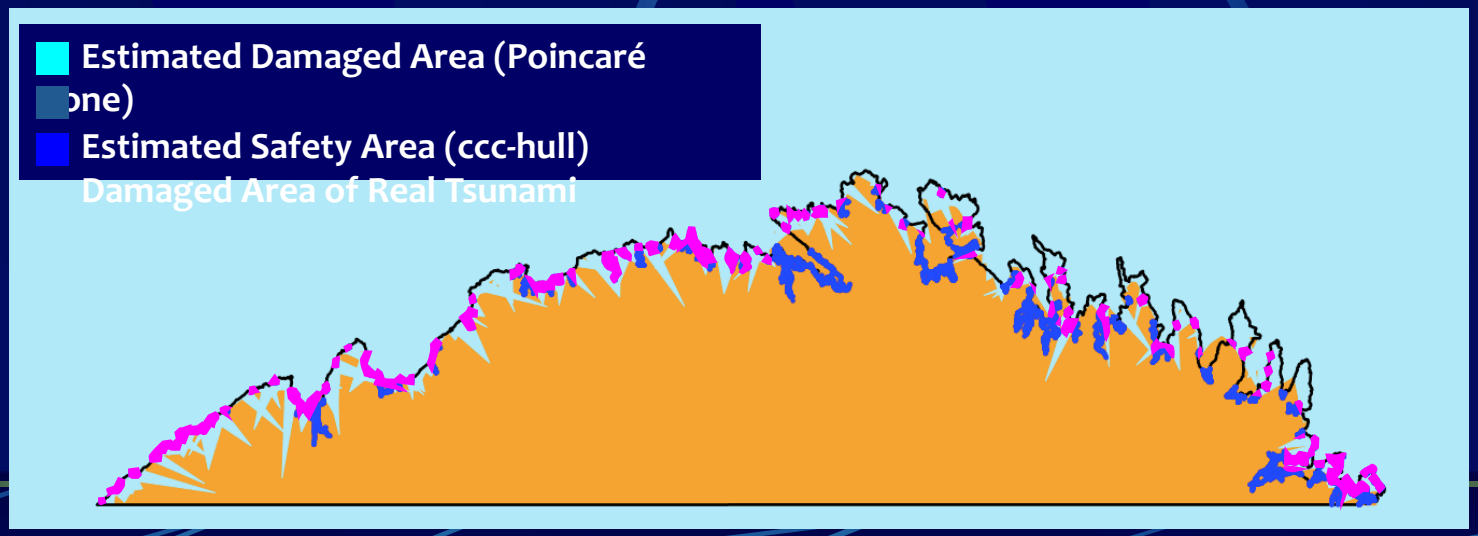


Result

- Ccc-hull for NHPP points using elevations ($\rho = 0.13\pi$)



- Ccc-hull for NHPP points by our algorithm ($\rho = 0.03\pi$)



まとめ

- リスクの評価と確率モデルとその必要性
- 災害と極値分布
- 線分の方向のモデル
- 土砂災害と楕円近似
- 津波の被害評価の総合指標とポアンカレコーン