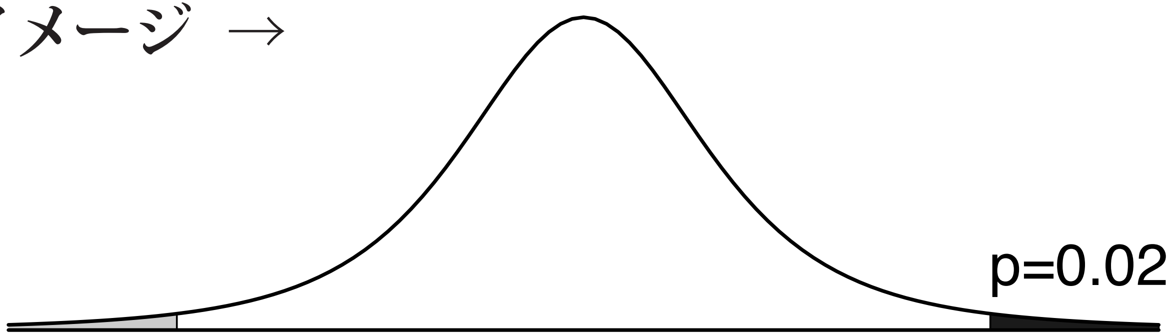
# 極値統計解析を行う上で生じる考え方の様々な相違（極値解析の使用上の誤解，極値理論に対する心得違い）

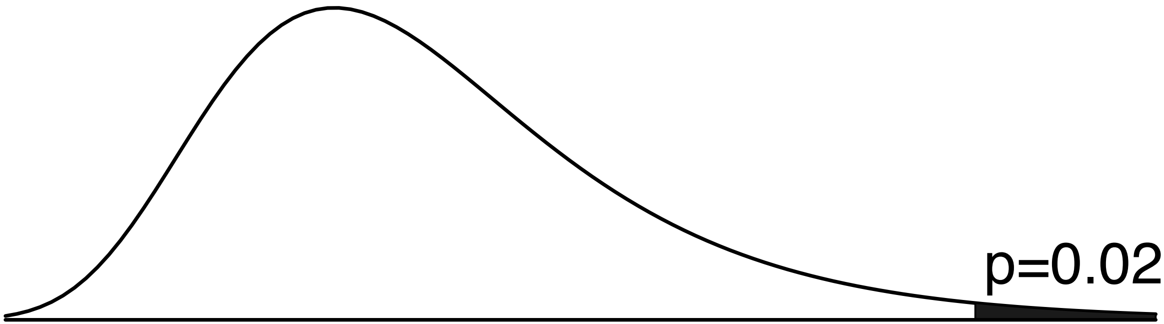
我々は確率・統計の世界でさまよう受難の民なのか？

名古屋工業大学・北野利一

大学の初年度に習う統計学のイメージ →




↓ 極値統計学のイメージ



**考え方の相違１）**

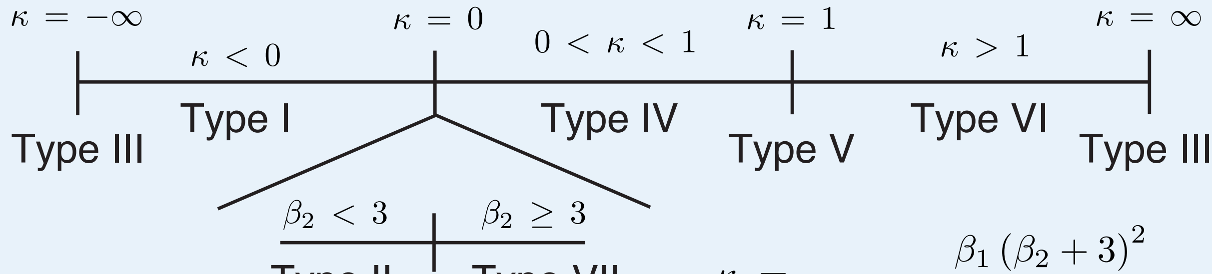
極値分布では無い分布を積極的に使う立場

大きな風呂敷を広げる，どのような風呂敷を？

## ピアソン分布族（ピアソン分布システム）

$$\frac{d\log f}{dy} = \frac{y+a}{b_0 + b_1 y + b_2 y^2 + \cdots}, \qquad \mu'_n = \int y^n f\, dy$$

$$a\mu'_n + nb_0\mu'_{n-1} + (n+1)b_1\mu'_n + (n+2)b_2\mu'_{n+1} + \cdots = -\mu'_{n+1}$$




$$\kappa = \frac{\beta_1 (\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)}$$

$$\sqrt{\beta_1} = \mu_3/\mu_2^{3/2}, \quad \beta_2 = \mu_4/\mu_2^2, \qquad \mu_n = \int (y-\mu_1)^n f\, dy$$

# 考え方の相違１）

極値分布では無い分布を積極的に使う立場

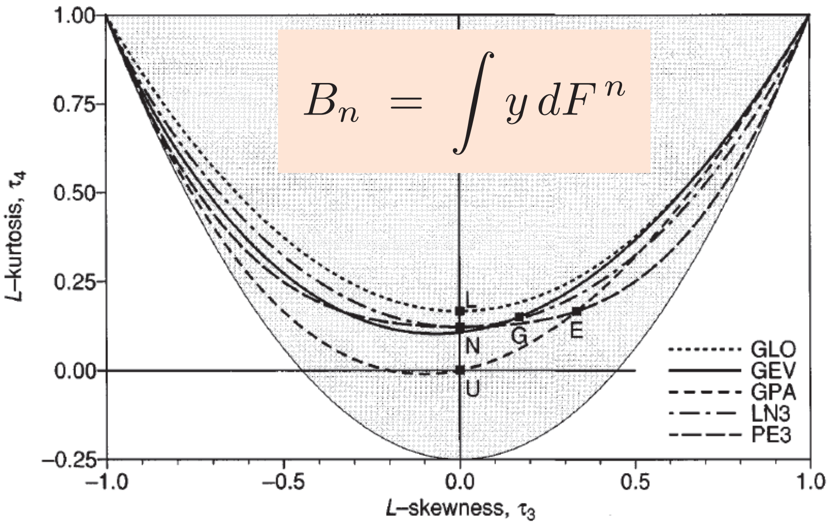
大きな風呂敷を広げる，どのような風呂敷を？

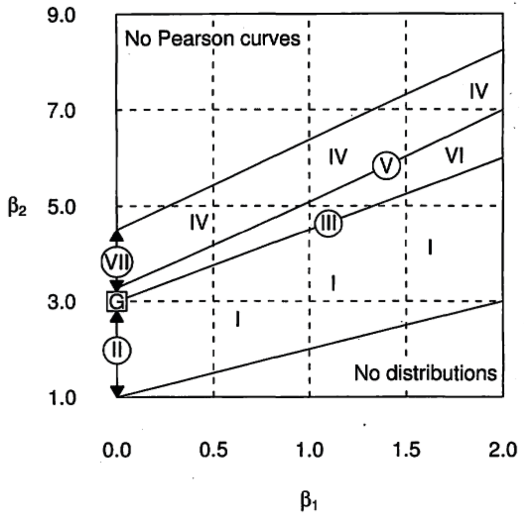
ピアソン分布族（ピアソン分布システム）

$$\frac{d\log f}{dy} = \frac{y+a}{b_0 + b_1 y + b_2 y^2 + \cdots},$$

$$\mu'_n = \int y^n f\,dy = \int y^n\,dF$$

$$B_n = \int y\,dF^n$$



**考え方の相違１）**

極値分布を使う立場

極値統計解析であるから，極値分布を使う.

**我々が欲しいのは，年最大値分布ではない.**

**再現期間における最大値分布が欲しい.**

**R 年確率外力とは，その代表値である.**

**それなのに, なぜ, 年最大値分布を考えるのか？**

年最大値分布と R 年最大値分布の関係

１．確率分布として，本来的に，関係が成立する条件

２．推定するために，論理的に成立すべき条件

３．推定に伴う誤差を考えて，データで確認できる限界

再現期間の定義１：年最大値分布の超過確率の逆数で定義

$$q_1 = 1 - F_1(y_R) = \frac{1}{R}$$

再現期間の定義２（提案）：生起率の逆数で定義

$$\lambda_1 = \frac{1}{R} \qquad \text{cf.} \ f = \frac{1}{T}$$

定義２を使うメリット：
１）そもそも定義が，単純明快！（周期 - 周波数と同じ関係）
２）GEV の表現で R 年最大値分布の位置母数が再現レベルになる
３）ガンベル確率紙の横軸を再現期間の対数で表現
（→図を見る時に，単純明快）

定義１を使うデメリット：
４）ブロック（単位期間）に依存している（論理的に困る！）
５）年最大値分布に対して，
再現期間が１年である確率外力が定義できない（困る！）
６）ガンベル確率紙などにおいて，
基準化変量（の値）に馴染めない（わかりにくい）

生起率と累積確率の関係：

$$F_1(y_R) = e^{-\lambda_1(y_R)}$$

（後に，その背景を示す）

■ ２年を１ブロックでみる時，
１年を１ブロックでみた時の再現期間と整合するか？

定義１では，残念ながら整合しない

$$F_2(y_R) = 1 - \frac{1}{R/2} = 1 - \frac{2}{R}$$

$$F_2(y_R) = F_1^2(y_R) = \left(1 - \frac{1}{R}\right)^2 = 1 - \frac{2}{R} + \frac{1}{R^2}$$

定義２では，キッカリ整合する

$$F_1(y_R) = e^{-\lambda_1(y_R)}$$

$$\rightarrow \quad F_2(y_R) = F_1^2(y_R) = e^{-2\lambda_1(y_R)} = e^{-\lambda_2(y_R)}$$

$$\rightarrow \quad \lambda_2(y_R) = 2\lambda_1(y_R) = \frac{2}{R} = \frac{1}{R/2}$$

待ち時間の超過確率 $\rightarrow$ $t > 1$ で生起する確率
$\rightarrow$ $0 \leq t \leq 1$ で生起しない確率

$$1 - F_{\mathrm{Exp}}(t=1) = f_{\mathrm{Poisson}}(k=0)$$

単位期間（1年）に生起数ゼロ
$\rightarrow$ 生起しない確率

$$= \exp(-\lambda_1)\Big|_{\lambda_1=\lambda_1(y)} = F_1(y)$$

外力 $y$ の累積確率
$\rightarrow$ $y$ を超える外力が生起しない確率

生起率と外力の大きさを関係づける式

$$\lambda_1(y) = \left(1 + \xi\,\frac{y - \mu_1}{\sigma_1}\right)^{-1/\xi}$$

■ 期間 n 年における生起率を次式の関数とすれば, その期間最大値分布が GEV となる.

$$\lambda_n(y) = \left(1 + \xi\,\frac{y - \mu_n}{\sigma_n}\right)^{-1/\xi} \quad \leftrightarrow \quad y = \mu_1 + \sigma_1\,\frac{\lambda_1^{-\xi} - 1}{\xi}$$

$$\left(= \mu_1 + \sigma_1 \log\frac{1}{\lambda_1}\right) \ (\text{for } \ \xi = 0)$$

$\lambda_R = R\lambda_1$ に対して，形を変えない（相似形を与えている！）.

$$y = \mu_1 + \sigma_1\,\frac{R^{\xi}\lambda_R^{-\xi} - R^{\xi} + R^{\xi} - 1}{\xi}$$

$$= \underbrace{\mu_1 + \sigma_1\,\frac{R^{\xi} - 1}{\xi}}_{= \mu_R} + \underbrace{R^{\xi}\sigma_1}_{= \sigma_R}\frac{\lambda_R^{-\xi} - 1}{\xi} = \mu_R + \sigma_R\,\frac{\lambda_R^{-\xi} - 1}{\xi}$$

この時，$\lambda_R(\mu_R) = R\lambda_1(\mu_R) = 1$ となるので,

定義 2 の意味で，再現期間 R 年の確率外力は，$\mu_R$ となる（単純明快）.

また, この時,（後に示す）経験度 K を用いて, その推定誤差分散は次式となる.

$$V(\hat{\mu}_R) = \frac{\sigma_R^2}{K} \qquad \text{cf.}\ \ V(\bar{X}_n) = \frac{\sigma^2}{n}$$

（これも単純明快！）

**考え方の相違１）**

極値分布を使う立場

極値統計解析であるから，極値分布を使う.

**我々が欲しいのは，年最大値分布ではない.**
**再現期間における最大値分布が欲しい.**
**R 年確率外力とは，その代表値である.**

**それなのに, なぜ, 年最大値分布を考えるのか？**

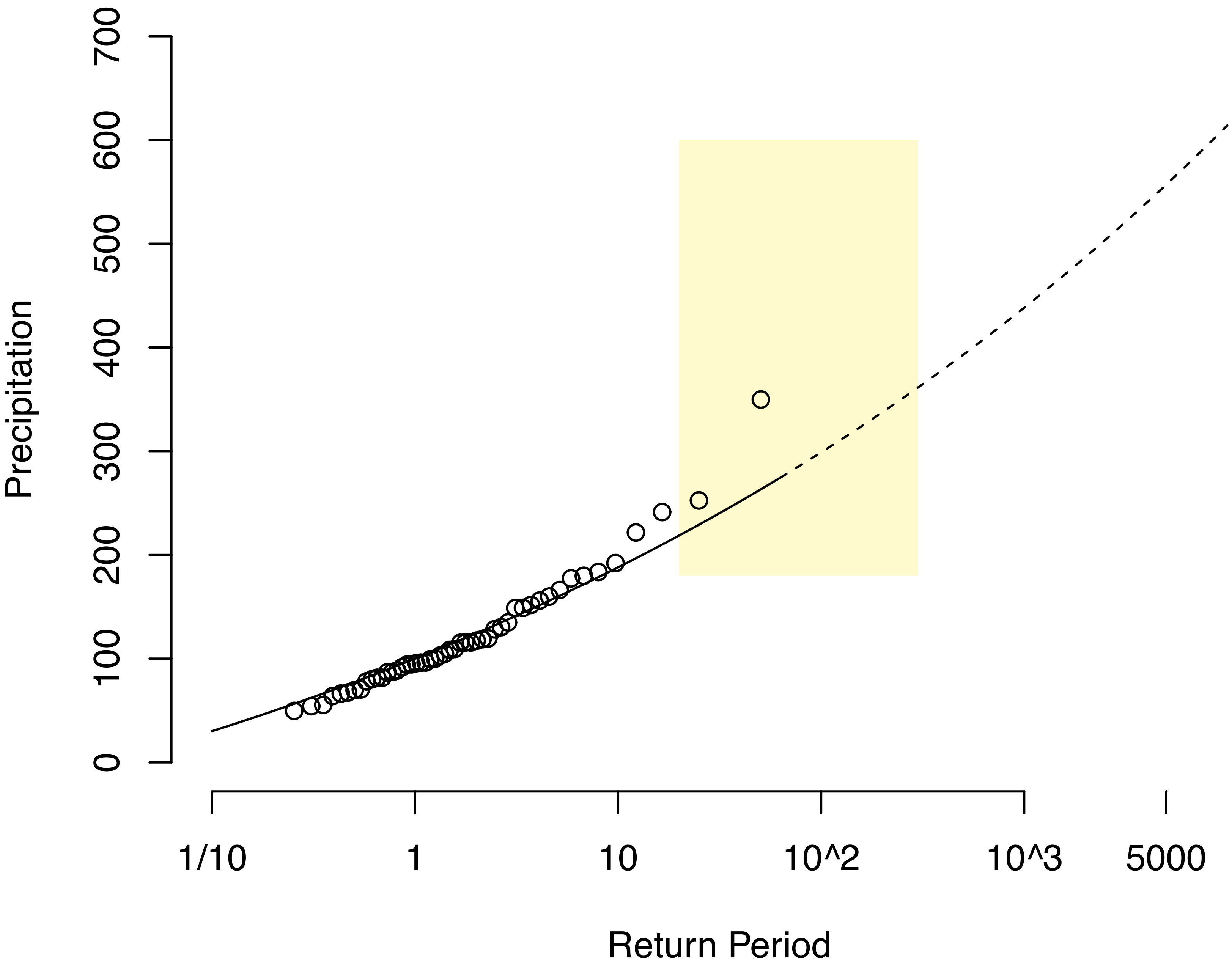
年最大値分布と R 年最大値分布の関係

**１．確率分布として，本来的に，関係が成立する条件**

２．推定するために，論理的に成立すべき条件

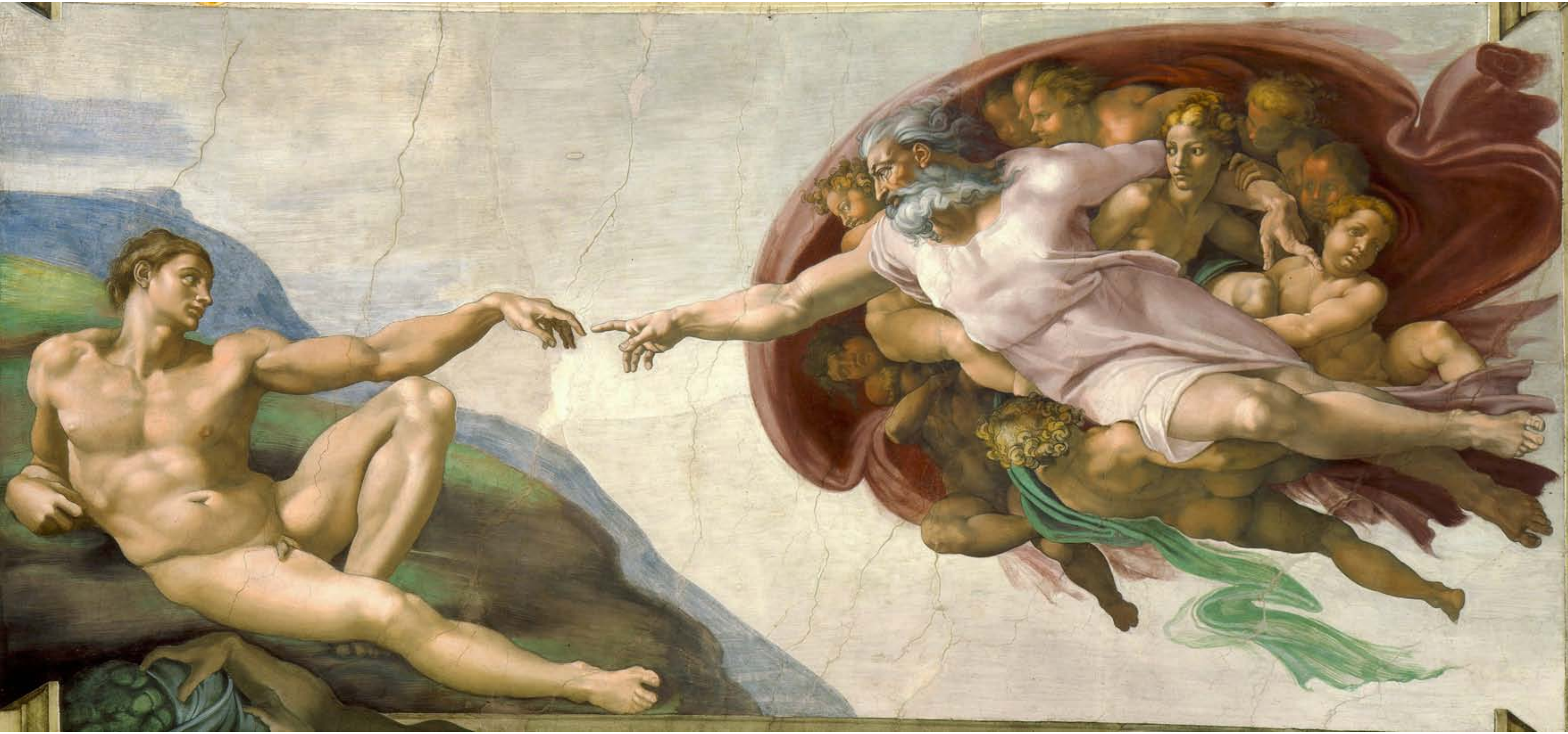
３．推定に伴う誤差を考えて，データで確認できる限界

# 外挿できる根拠



700
600
500
400
300
200
100
0
Precipitation
1/10
1
10
10^2
10^3
5000
Return Period

# 地上と天空をつなぐもの

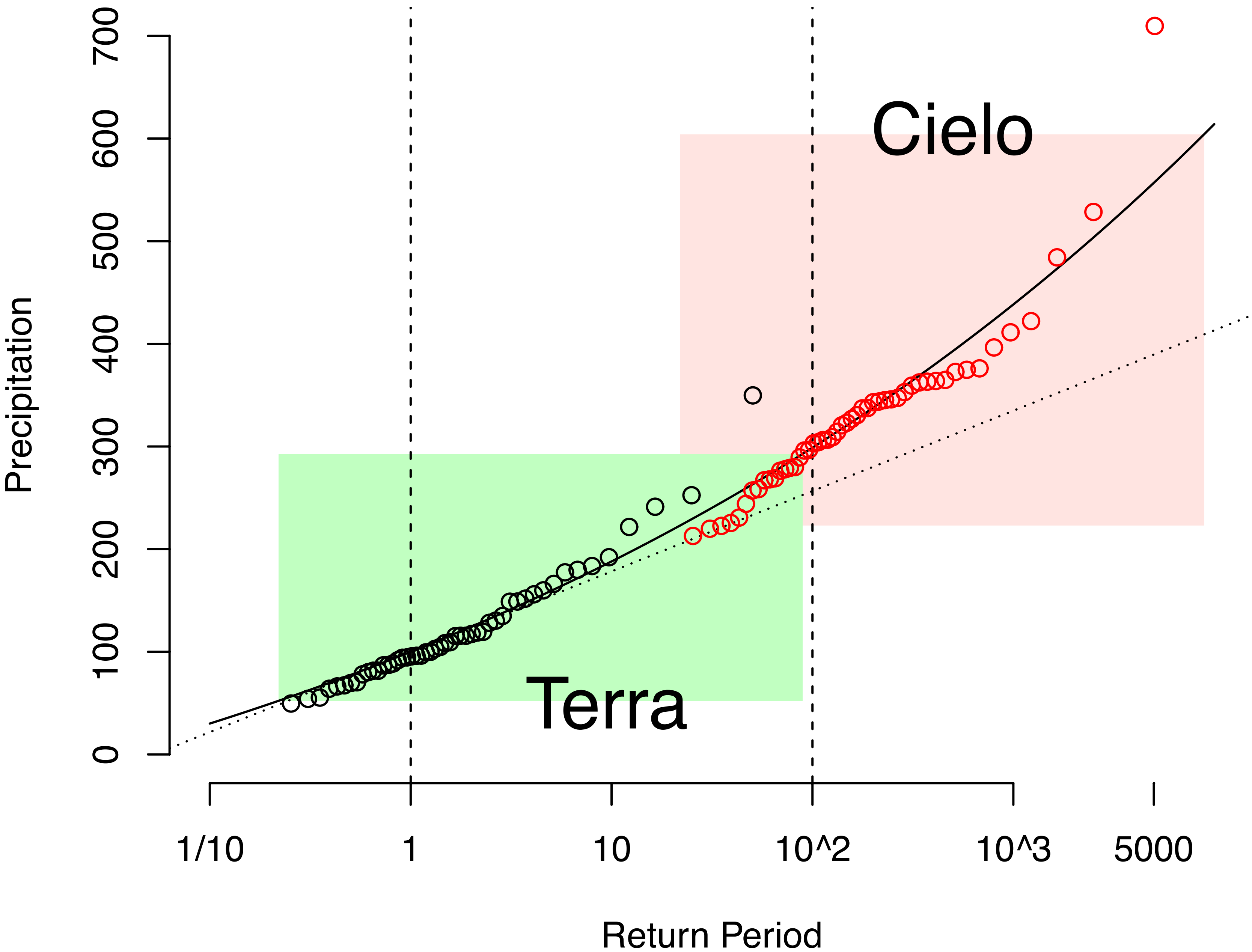


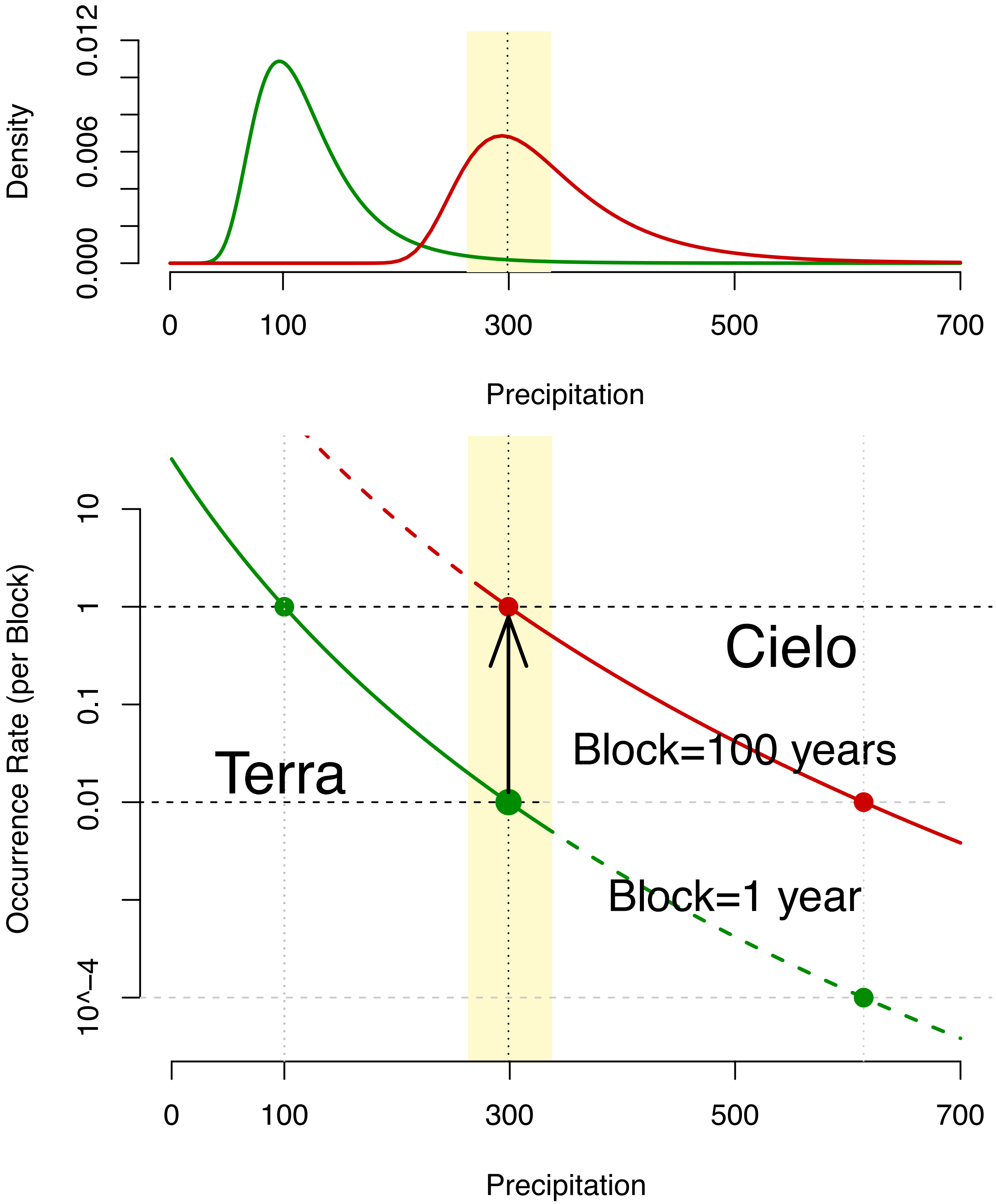
システィーナ礼拝堂 500 年祭記念
ミケランジェロ展／国立西洋美術館 2013 年 9 月 6 日～ 11 月 17 日

# 地上と天空をつなぐもの


システィーナ礼拝堂 500 年祭記念
ミケランジェロ展／国立西洋美術館 2013 年 9 月 6 日～ 11 月 17 日

# 地上と天空をつなぐもの






重要なことは，頻度の比例関係（相似性）が成立することである．

$$\lambda_1 : 1 = \lambda_n : n$$

すなわち，

$$\lambda_n = n\lambda_1$$

が成立すること

例，

$$\lambda_1 = 1/100 = 0.01$$

$$\lambda_n = 100\,\lambda_1 = 1$$

proportionality

$$\lambda_1 : 1 = \lambda_n : n \quad \to \quad \lambda_n = n\lambda_1$$

$$\to \quad \log\lambda_n = \log n + \log\lambda_1 \quad \to \quad d\log\lambda_n = d\log\lambda_1$$

de Haan

$$-\frac{d}{d\log\lambda}\log\left(-\frac{dy}{d\log\lambda}\right) = \xi$$

$$\frac{1-F(y)}{f(y)} \approx \frac{-\lambda(y)}{d\lambda(y)/dy} = -\frac{dy}{d\log\lambda}$$

$$\frac{d}{dy}\left\{\frac{1-F(y)}{f(y)}\right\} \approx \frac{-dy}{d\log\lambda}\left\{\frac{d}{dy}\left(-\frac{dy}{d\log\lambda}\right)\right\}\Big/\left(\frac{-dy}{d\log\lambda}\right) = \frac{-d}{d\log\lambda}\log\left(-\frac{dy}{d\log\lambda}\right)$$

von Mises

$$\lim_{y\to y_b}\frac{d}{dy}\left(\frac{1-F}{f}\right) = \xi$$

Intensity measure

$$\lambda_i = \lambda(y;\boldsymbol{\theta}_i) = \exp\left\{-\frac{1}{\xi}\log\left(1+\xi\,\frac{y-\mu_i}{\sigma_i}\right)\right\}$$

$$\boldsymbol{\theta}_i = \{\mu_i, \sigma_i, \xi\}$$

proportionality

**頻度の連続性**

★ $$\lambda_n = n\lambda_1 \Big|_{\lambda_1=\frac{1}{n}} = 1$$

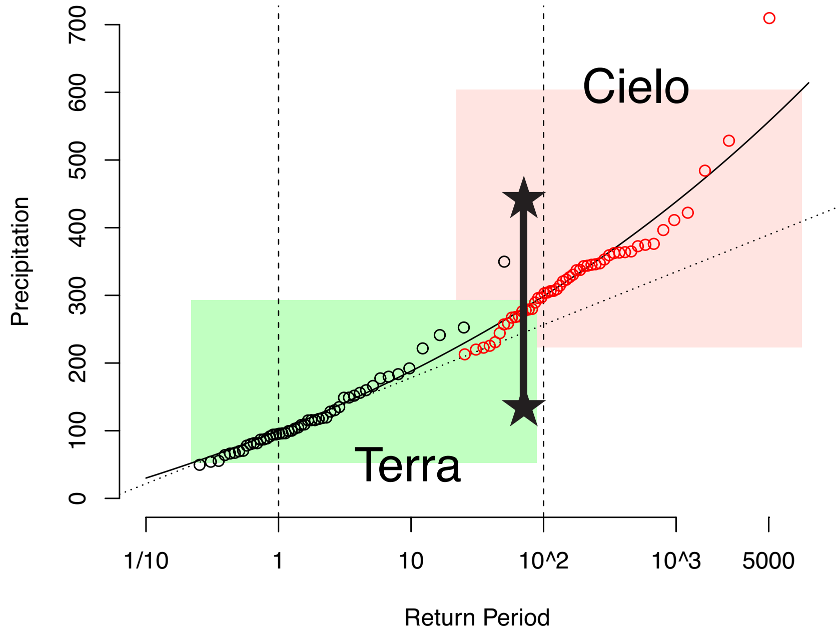
★ $$y\Big|_{\lambda_n=1} = y\Big|_{\lambda_1=\frac{1}{n}} \Leftrightarrow \mu_n = \mu_1 + \sigma_1 \frac{n^\xi - 1}{\xi}$$

**確率外力の連続性 ← 任意の分布関数でカーブフィッティング（極値解析）をする場合には，この関係しか考えていない．**

★ $$-\frac{dy}{d\log\lambda}\Big|_{\lambda_n=1} = -\frac{dy}{d\log\lambda}\Big|_{\lambda_1=\frac{1}{n}}$$

**尺度（スケール）の連続性**

$$\Leftrightarrow \quad \sigma_n = n^\xi \sigma_1$$




★ de Haan

**形状の連続性**

$$-\frac{d}{d\log\lambda}\log\left(-\frac{dy}{d\log\lambda}\right) = \xi$$

相似性

$$-\frac{dy}{d\log\lambda} = \frac{\sigma}{\lambda^\xi},$$

$$y = \mu + \sigma\frac{\exp(-\xi\log\lambda) - 1}{\xi}$$

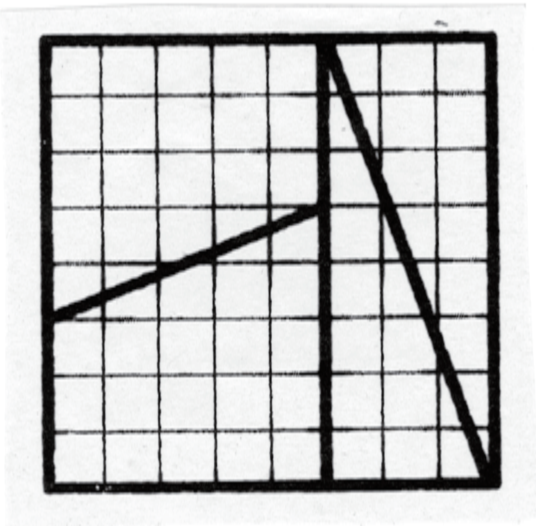
（$\mu$, $\sigma$ は積分定数）

Intensity measure

$$\lambda_i = \lambda(y;\boldsymbol{\theta}_i) = \exp\left\{-\frac{1}{\xi}\log\left(1+\xi\frac{y-\mu_i}{\sigma_i}\right)\right\}$$

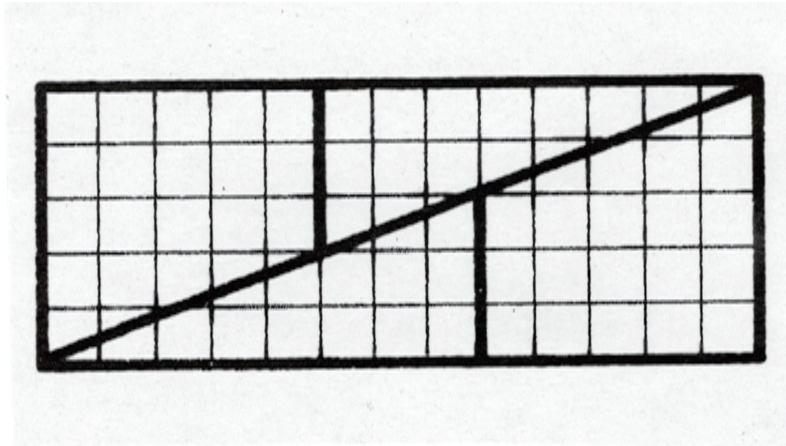
相似性，あるいは，比例関係（Proportionality）が崩れると困りますネ！

（我々人間の思考で，根源的に要求するものの１つ）



$\neq$

?



$$8 \times 8 \neq 13 \times 5$$

## 考え方の相違２）

数理統計学の立場，基本的には，最尤法を使う.
（最尤法でマズイ時に，最尤法以外を用いる）

**我々が欲しいのは，年最大値分布ではない.**
**再現期間における最大値分布が欲しい.**
**R 年確率外力とは，その代表値である.**

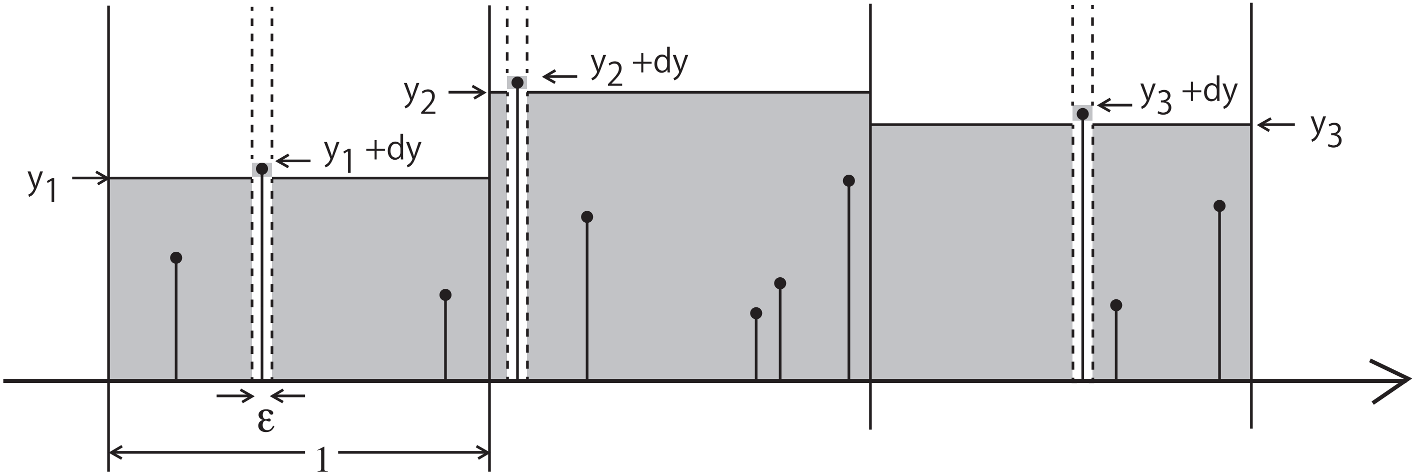
**それなのに, なぜ, 年最大値分布を考えるのか？**

年最大値分布と R 年最大値分布の関係

１．確率分布として，本来的に，関係が成立する条件

**２．推定するために，論理的に成立すべき条件**

３．推定に伴う誤差を考えて，データで確認できる限界

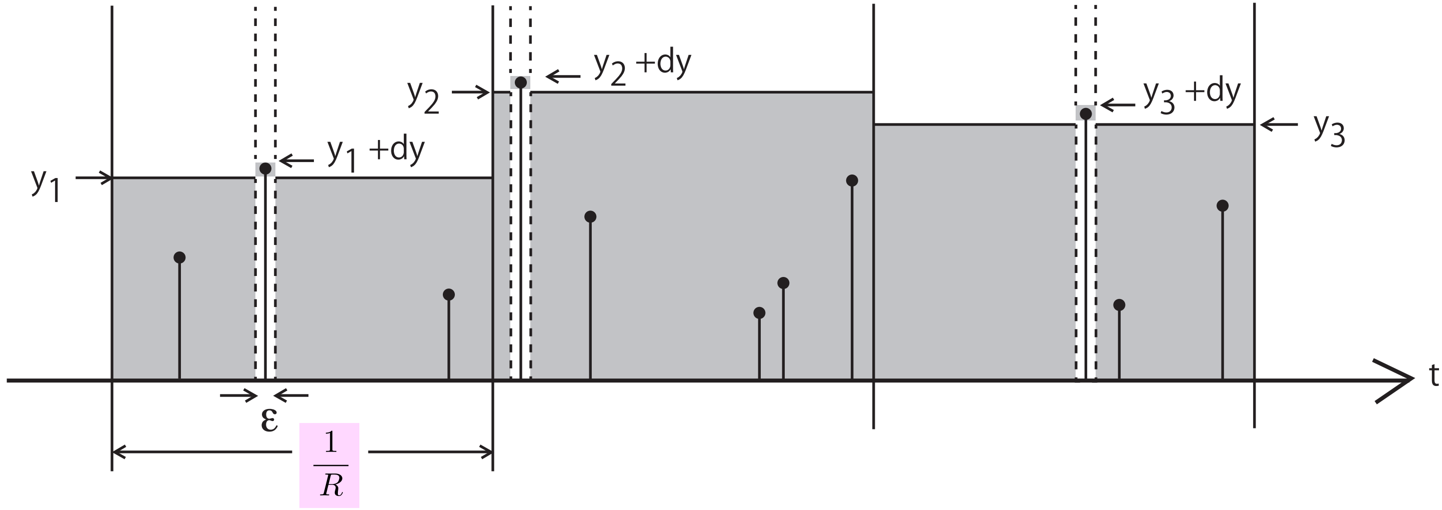



$$L_{-,i} = \frac{\lambda^k}{k!}e^{-\lambda}\bigg|_{k=0,\ \lambda=(1-\varepsilon)\lambda_1(y_i)} \approx \exp\Big\{-\lambda(y_i,\boldsymbol{\theta}_1)\Big\}$$

$$L_{+,i} = \frac{\lambda^k}{k!}e^{-\lambda}\bigg|_{k=1,\ \lambda=\varepsilon\lambda_1(y_i)} - \frac{\lambda^k}{k!}e^{-\lambda}\bigg|_{k=1,\ \lambda=\varepsilon\lambda_1(y_i+dy)} \approx -\varepsilon\, dy\, \frac{d\lambda}{dy}(y_i,\boldsymbol{\theta}_1)$$

$$L_i = L_{-,i}\,L_{+,i} \propto \exp\Big\{-\lambda(y_i,\boldsymbol{\theta}_1)\Big\}\left\{-\frac{d\lambda}{dy}(y_i,\boldsymbol{\theta}_1)\right\}$$

$$= \frac{d}{dy}\exp\Big\{-\lambda(y_i,\boldsymbol{\theta}_1)\Big\} = f(y_i,\boldsymbol{\theta}_1) \qquad (\text{for} \quad \varepsilon \to 0)$$




$$L_{-,i} \;=\; \frac{\lambda^k}{k!}e^{-\lambda}\bigg|_{k=0,\ \lambda=(\frac{1}{R}-\varepsilon)\lambda_R(y_i)} \;\approx\; \exp\left\{-\frac{\lambda(y_i,\boldsymbol{\theta}_R)}{R}\right\}$$

これは，R 年最大値分布の累積（確率）分布関数

$$L_{+,i} \;=\; \frac{\lambda^k}{k!}e^{-\lambda}\bigg|_{k=1,\ \lambda=\varepsilon\lambda_R(y_i)} - \frac{\lambda^k}{k!}e^{-\lambda}\bigg|_{k=1,\ \lambda=\varepsilon\lambda_R(y_i+dy)} \;\approx\; -\,\varepsilon\,dy\,\frac{d\lambda}{dy}(y_i,\boldsymbol{\theta}_R)$$

$$L_i \;=\; L_{-,i}\,L_{+,i} \;\propto\; \exp\left\{-\frac{\lambda(y_i,\boldsymbol{\theta}_R)}{R}\right\}\left\{-\frac{d\lambda}{dy}(y_i,\boldsymbol{\theta}_R)\right\}$$

$$\propto\; \boxed{\frac{d}{dy}\exp\left\{-\,\lambda(y_i,\boldsymbol{\theta}_1)\right\} \;=\; f(y_i,\boldsymbol{\theta}_1)}$$

年最大値分布の尤度に帰着するのである！

なぜなら，

$$\lambda_R \;=\; R\lambda_1$$

$$\leftrightarrow\quad \lambda(y_i,\boldsymbol{\theta}_1) \;=\; \frac{\lambda(y_i,\boldsymbol{\theta}_R)}{R}$$

## 考え方の相違１）

極値分布を使う立場

極値統計解析であるから，極値分布を使う．

**我々が欲しいのは，年最大値分布ではない．**
**再現期間における最大値分布が欲しい．**
**R 年確率外力とは，その代表値である．**

**それなのに, なぜ, 年最大値分布を考えるのか？**

年最大値分布と R 年最大値分布の関係

１．確率分布として，本来的に，関係が成立する条件

２．推定するために，論理的に成立すべき条件

**３．推定に伴う誤差を考えて，データで確認できる限界**

# 経験度 とは？ （手短に何であるか説明してください，としばしば依頼があるのだが）．

$$E\,(\delta\log\lambda)^2 \;=\; V\left\{\log\lambda(y;\hat{\boldsymbol{\theta}})\right\} \;=\; \frac{\nabla'_{\boldsymbol{\theta}}\lambda\;\mathcal{I}_{\boldsymbol{\theta}}^{-1}\;\nabla_{\boldsymbol{\theta}}\lambda}{\lambda^2} \;=\; \frac{1}{K}$$

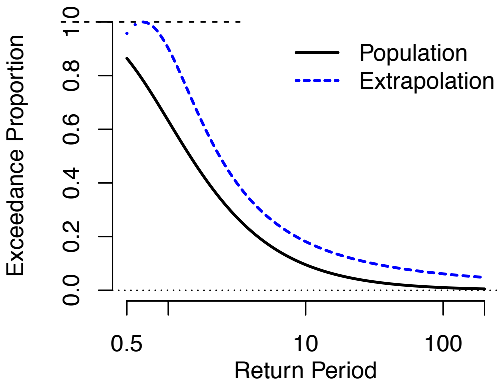
1）最も単純な経験度：

**外挿しないのなら，頻度の推定に用いるデータ数（そのもの）＝ 経験度**

（逆に言えば，**生起率関数を用いて，内部の情報を，外部に移転させることが，極値解析の効用の本質**であり，その**移転できる情報の量を表すのが経験度**と言える）

・ポアソン分布の生起率の推定：$\hat{\lambda} = \dfrac{k}{N}$，その誤差分散：$V(\hat{\lambda}) = \dfrac{\lambda}{N}$

$$\begin{aligned}
\frac{1}{K} &= V\left(\log\hat{\lambda}\right) \\
&= V(\hat{\lambda})\left(\frac{d\log\lambda}{d\lambda}\right)^2 \\
&= \frac{1}{N\lambda} \sim \frac{1}{k} \\
&\rightarrow \quad K = k
\end{aligned}$$



## 経験度 とは？ （手短に何であるか説明してください，としばしば依頼があるのだが）.

$$E\,(\delta \log \lambda)^2 \;=\; V\left\{\log \lambda(y;\hat{\boldsymbol{\theta}})\right\} \;=\; \frac{\nabla'_{\boldsymbol{\theta}}\lambda\ \mathcal{I}_{\boldsymbol{\theta}}^{-1}\ \nabla_{\boldsymbol{\theta}}\lambda}{\lambda^2} \;=\; \frac{1}{K}$$

1）最も単純な経験度：

**外挿しないのなら，頻度の推定に用いるデータ数（そのもの）＝ 経験度**

（逆に言えば，**生起率関数を用いて，内部の情報を，外部に移転させることが，極値解析の効用の本質**であり，その**移転できる情報の量を表すのが経験度**と言える）

・対数尤度による情報からも

$$f(k) \;=\; \frac{(N\lambda)^k}{k!}\, e^{-N\lambda} \quad\rightarrow\quad \ell \;=\; k\log\lambda \;-\; N\lambda \;+\; \log(N^k/k!)$$

$$-\frac{d\ell}{d\lambda} = N - \frac{k}{\lambda}, \quad \mathcal{I} = -\frac{d^2\ell}{d\lambda^2} = \frac{k}{\lambda^2} \quad\rightarrow\quad \frac{1}{K} = \frac{\mathcal{I}^{-1}}{\lambda^2} = \frac{1}{k}$$

$$\rightarrow\quad K \;=\; k$$

経験度 とは？ （手短に何であるか説明してください，としばしば依頼があるのだが）.

２）ベイズ的解釈　　生起率に対する推定誤差の集中の度合い

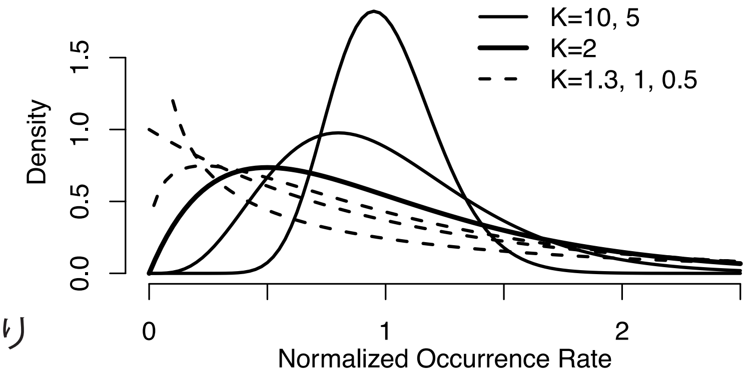
**～ ポアソン分布の自然共役となるガンマ分布の形状母数 ＝ 経験度**

比例関係の成立　$1 : R = K : L \quad \rightarrow \quad \frac{1}{R} = \frac{K}{L} = \lambda$

$$E(\hat{\lambda}) = \frac{K}{L}, \quad V(\hat{\lambda}) = \frac{K}{L^2},$$

$$f(\hat{\lambda}) = \frac{L^K}{\Gamma(K)} \hat{\lambda}^{K-1} \exp(-L\hat{\lambda})$$

$$V(\log \hat{\lambda}) = \frac{K/L^2}{(K/L)^2} = \frac{1}{K}$$




生起率関数を介して，ガンマ分布より
確率外力 $\mu_R$ の信頼区間を算定可能.

経験度が増大すれば，ガンマ分布 → 正規分布となる.
その時には，正規分布の分散に $V(\hat{\mu}_R) = \sigma_R^2/K$ を用いることができる.
逆に言えば，経験度が５程度以下であれば，誤差分布に正規分布が使えない.

外挿の限界：“２度あることは３回ある”（度と回の違いに注意！）

# 外挿の限界：

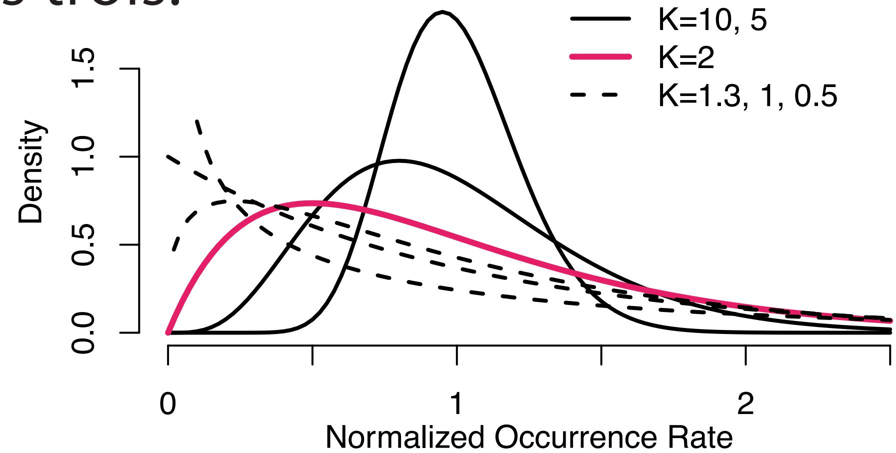
“２度あることは３回ある”

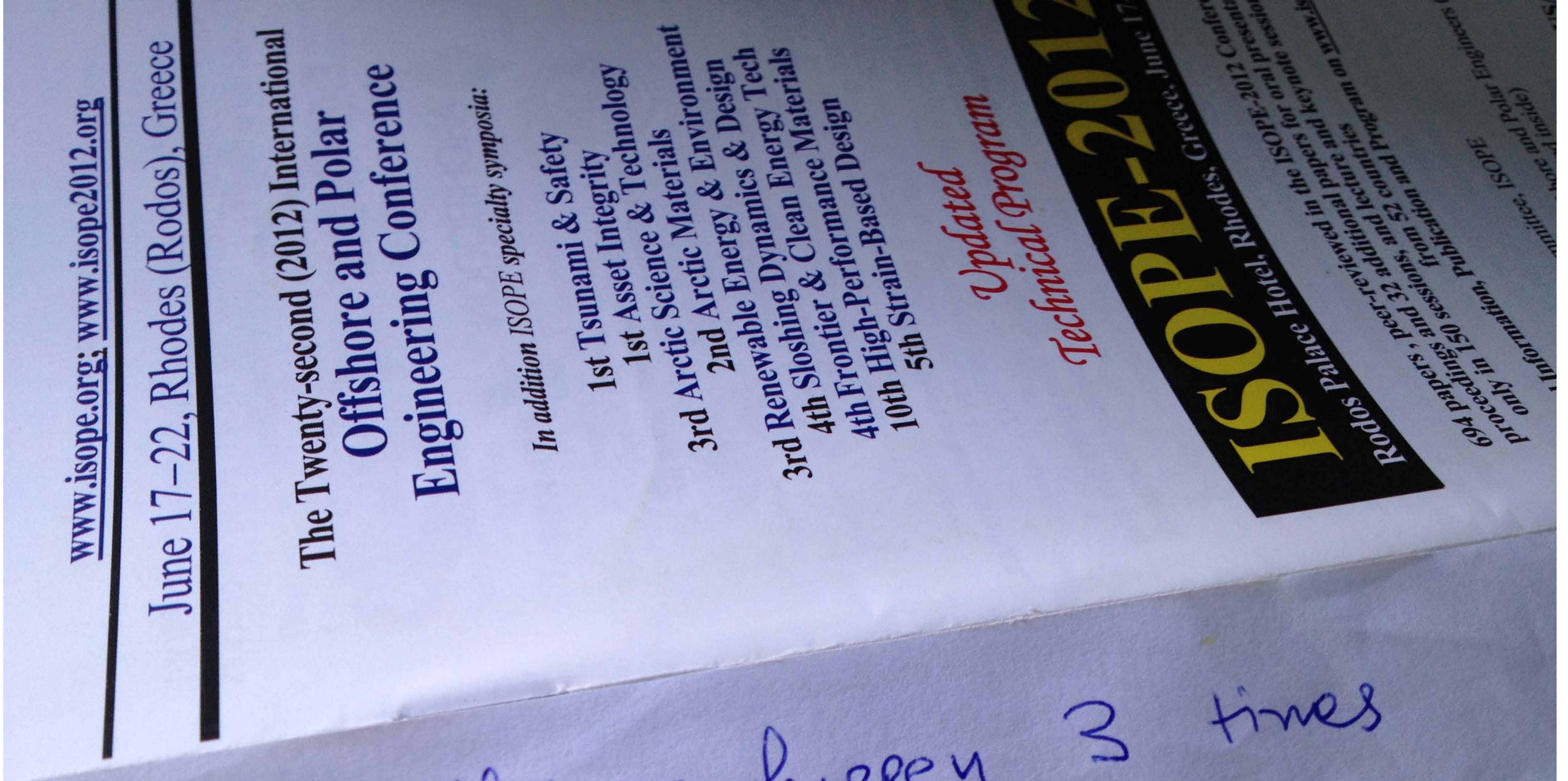
What happend twice will happen three times.

Non c'è due senza tre.

Non hay dos sin tres.

Jamais deus sans trois.

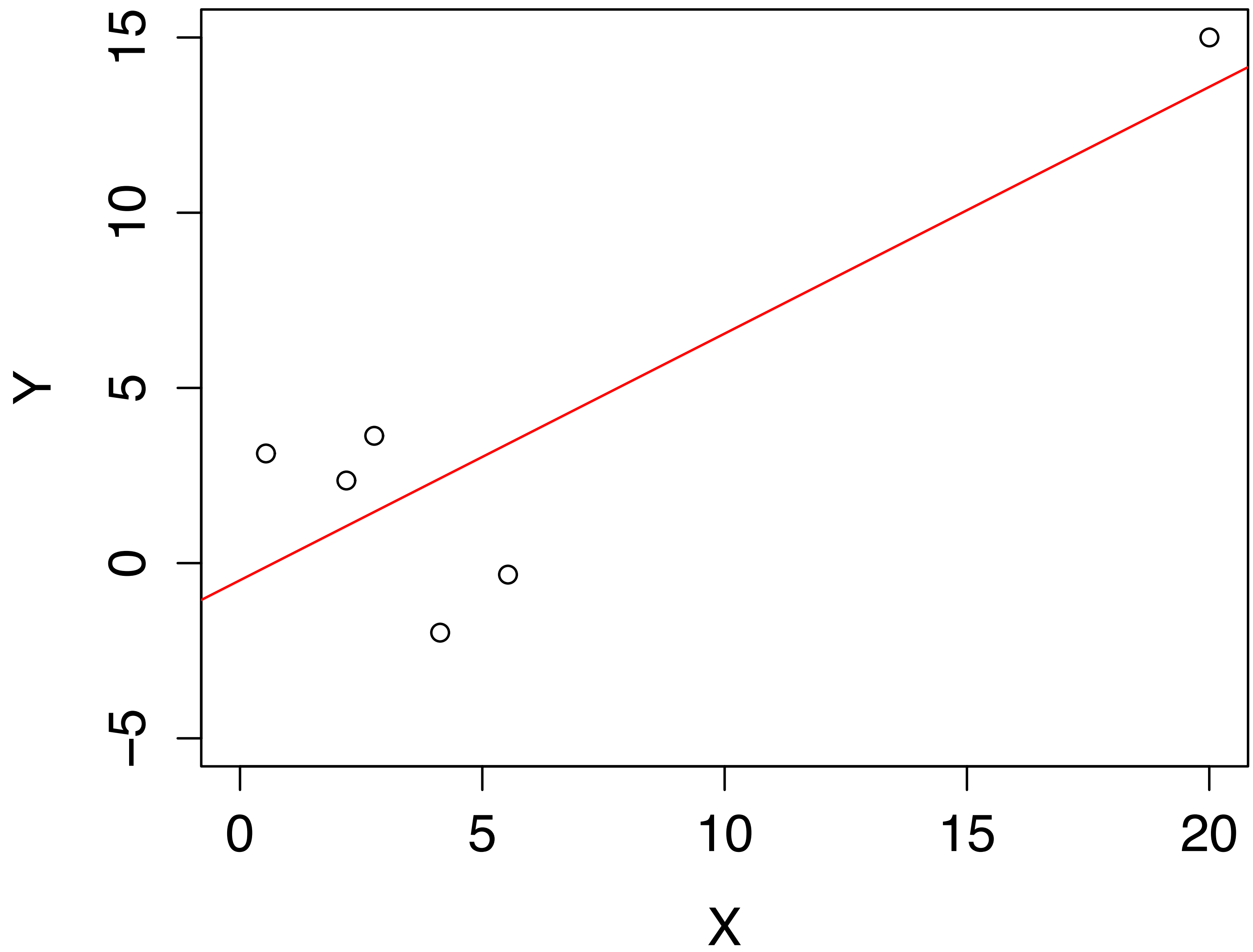




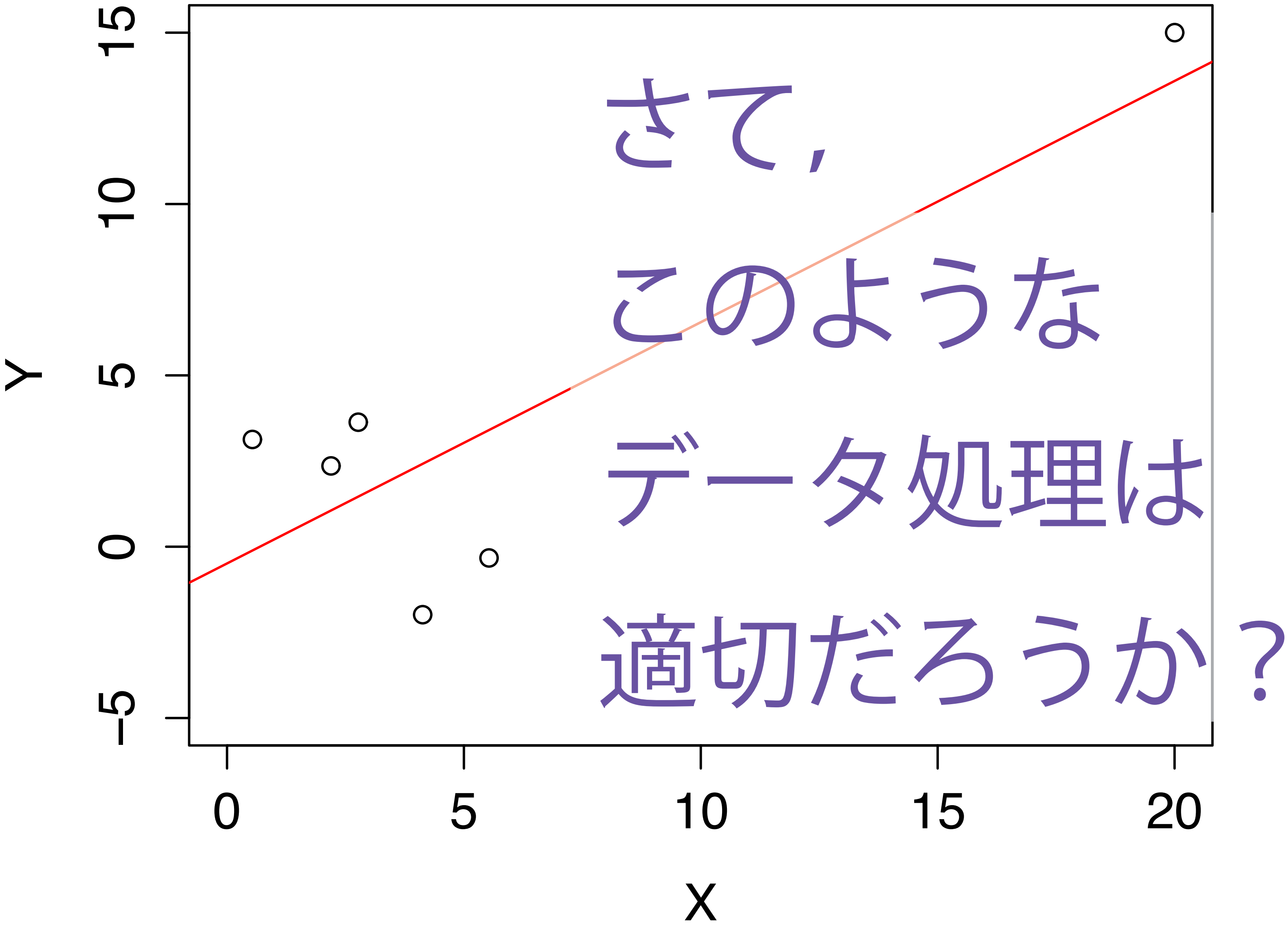

Bad things happen 3 times

Το κακό τρίγωνα.

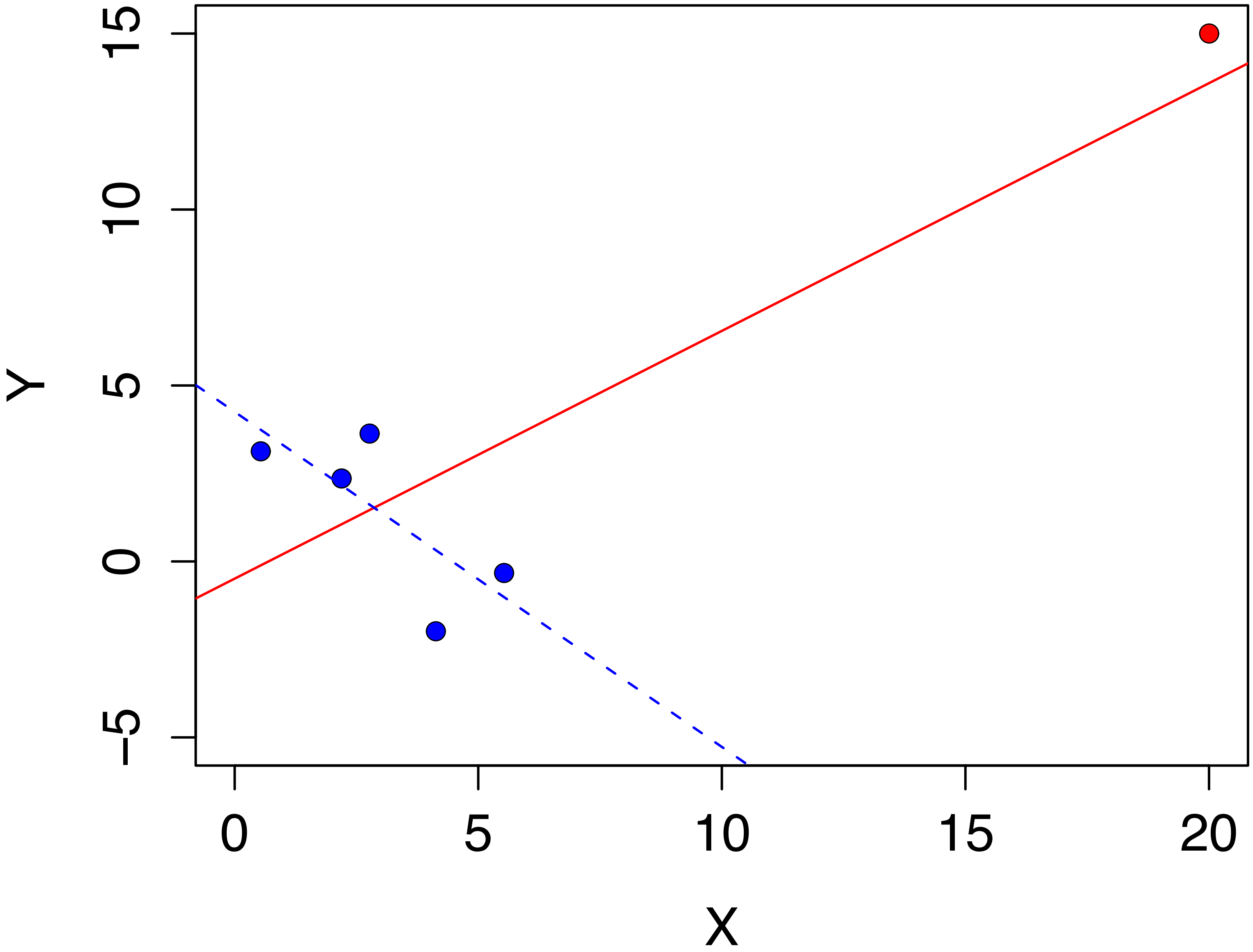
# いわゆる回帰分析の問題



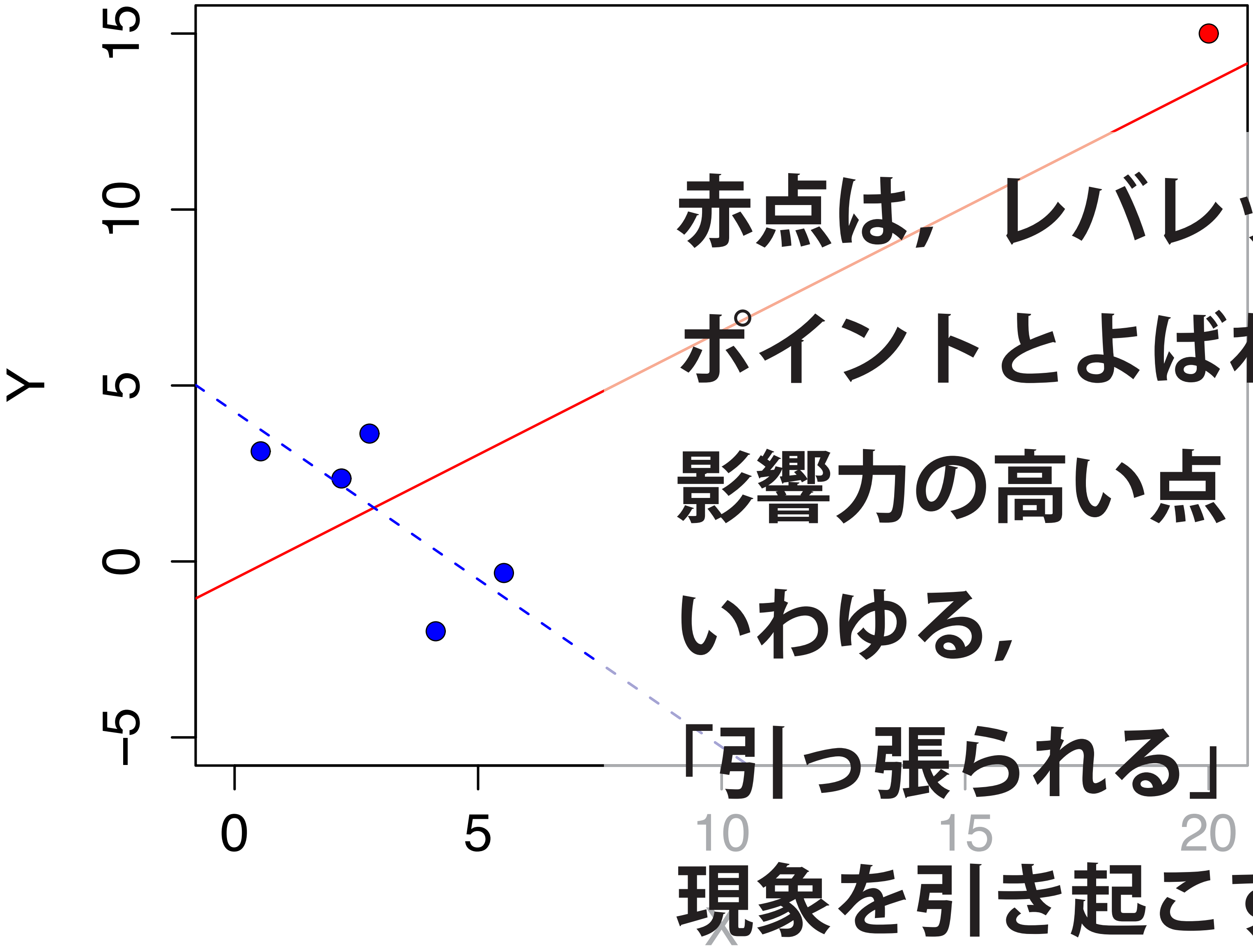
# いわゆる回帰分析の問題



# いわゆる回帰分析の問題

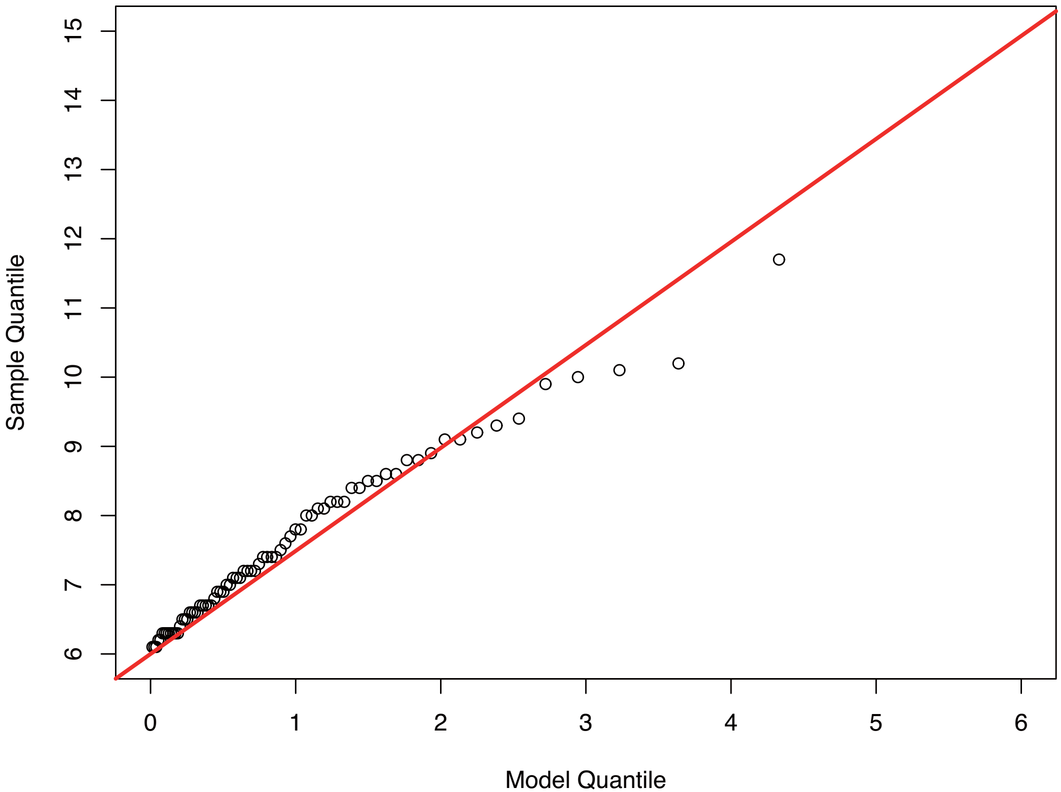


# いわゆる回帰分析の問題

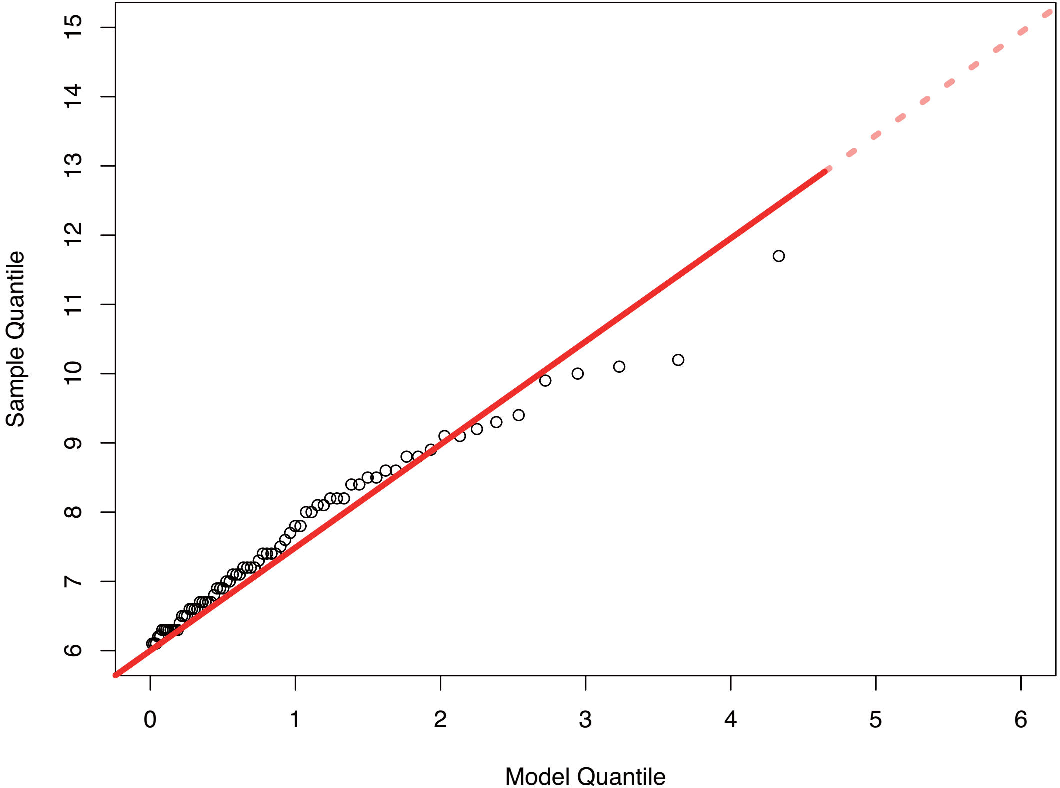



赤点は，レバレッジポイントとよばれる，影響力の高い点いわゆる，「引っ張られる」という現象を引き起こす．

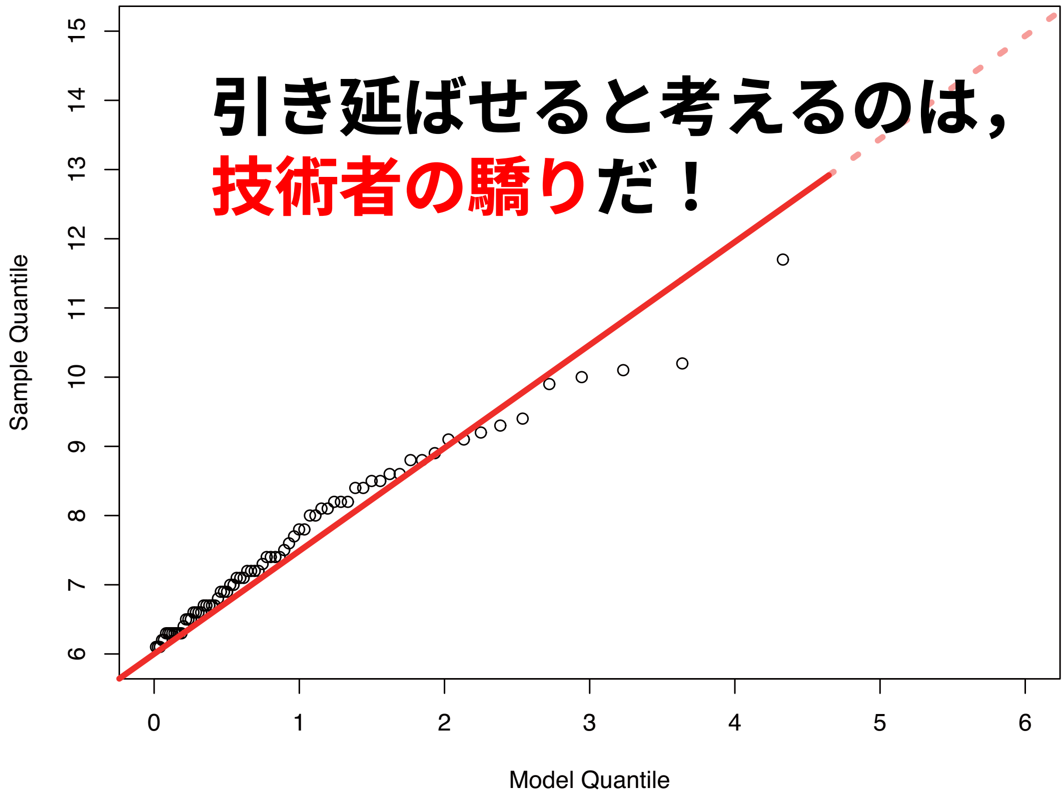
# 極値解析といえば，この図ですが，...



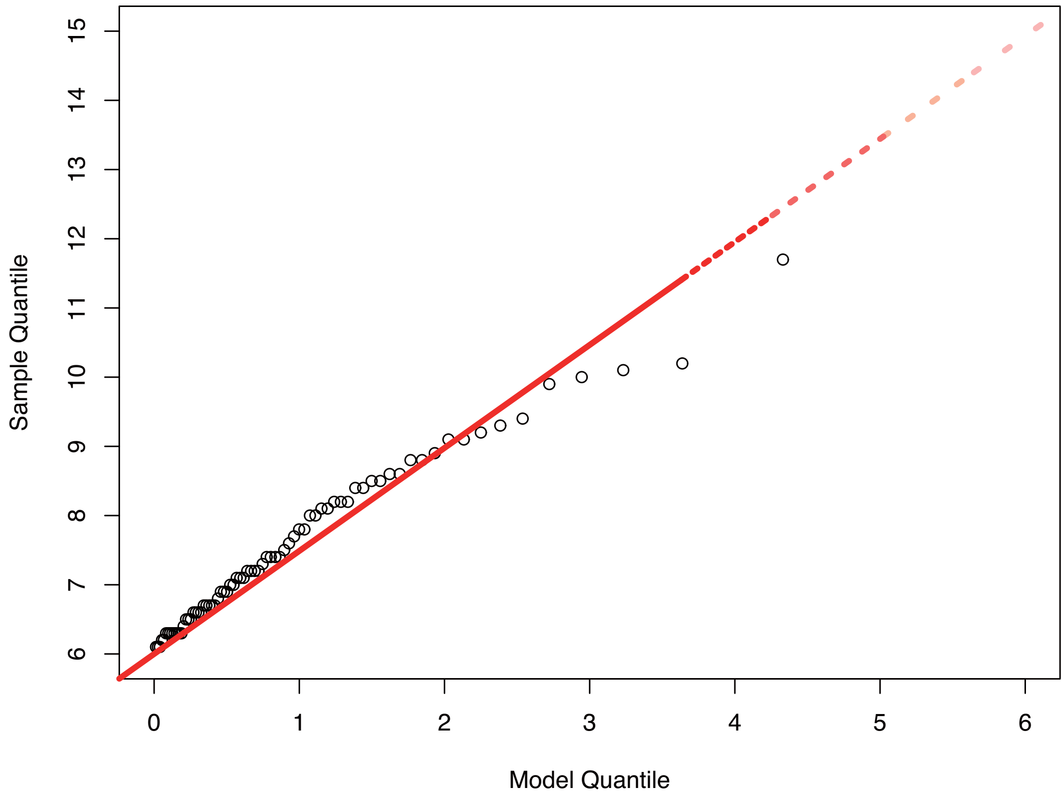
# 技術者の感覚 〜 どこまでも引き延ばして良いワケは無い（外挿の問題）



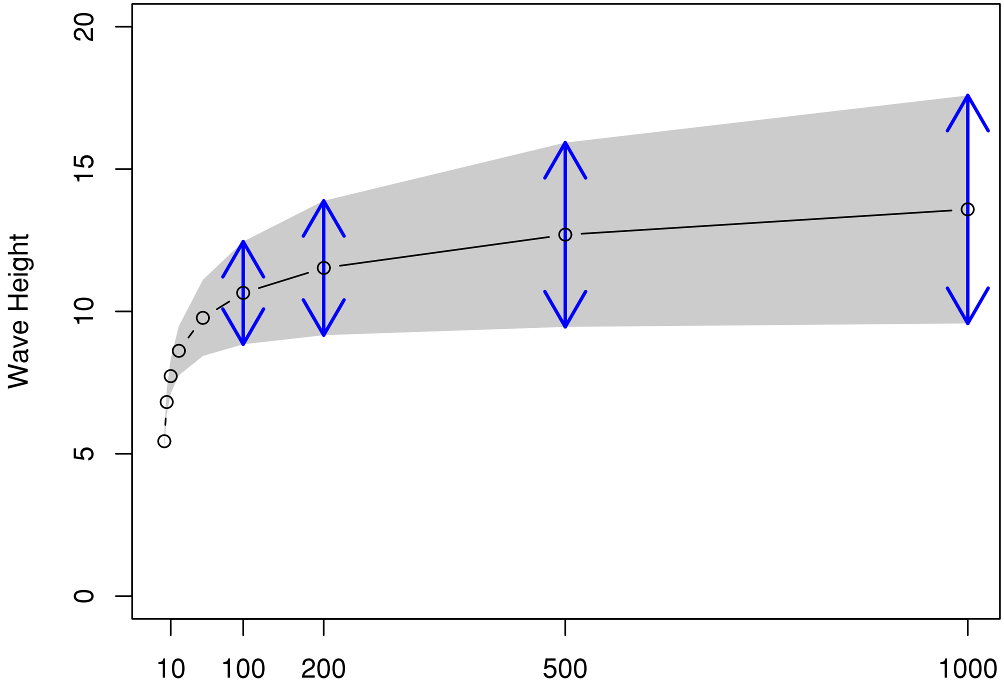
推定誤差がその原因.信頼区間という概念は相対的であり，役に立たない（ことに気付くべきだ）.



市民感覚でも，このように見えるのが，現実的．
問題は，1）濃淡の算出法，2）その打ち切り判断

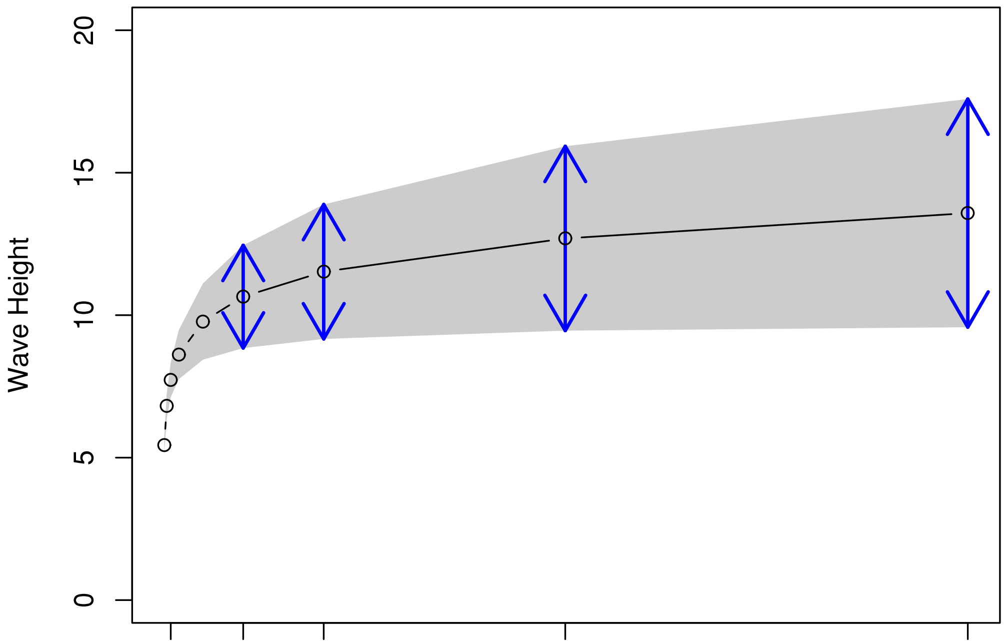


横軸に再現期間そのもので表示して，　...
信頼区間を描いて見ても,...　　　　（従来の議論）



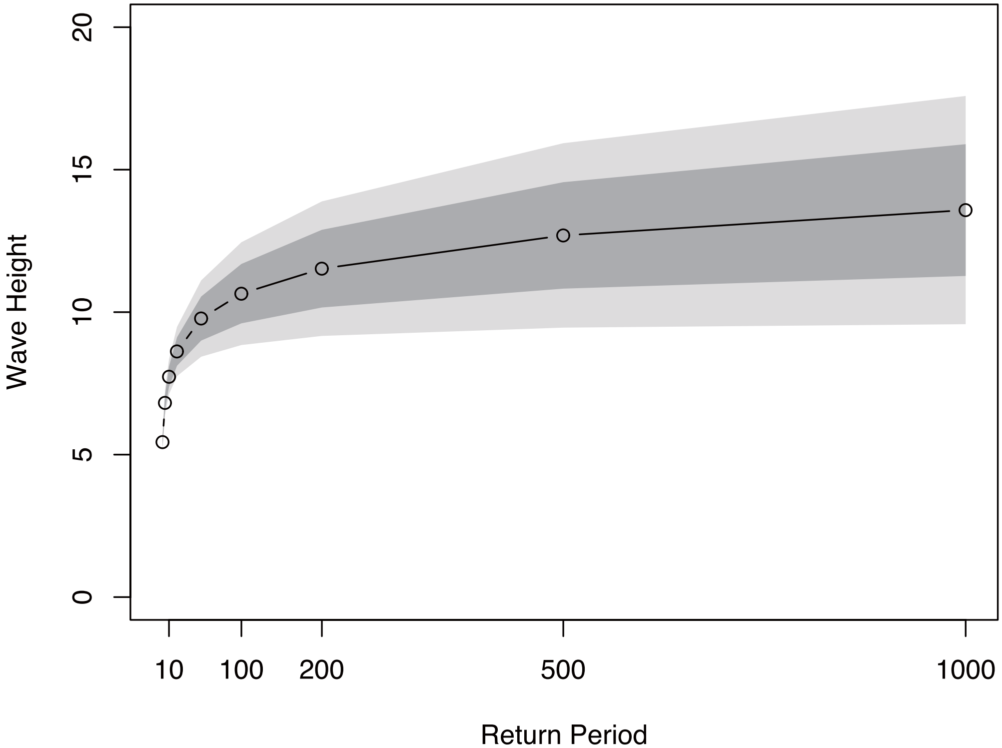

実のところ，良く分かりません.

むしろ，統計誤差や信頼区間という概念について，
過小評価（大したこと無い）を助長するだけです．

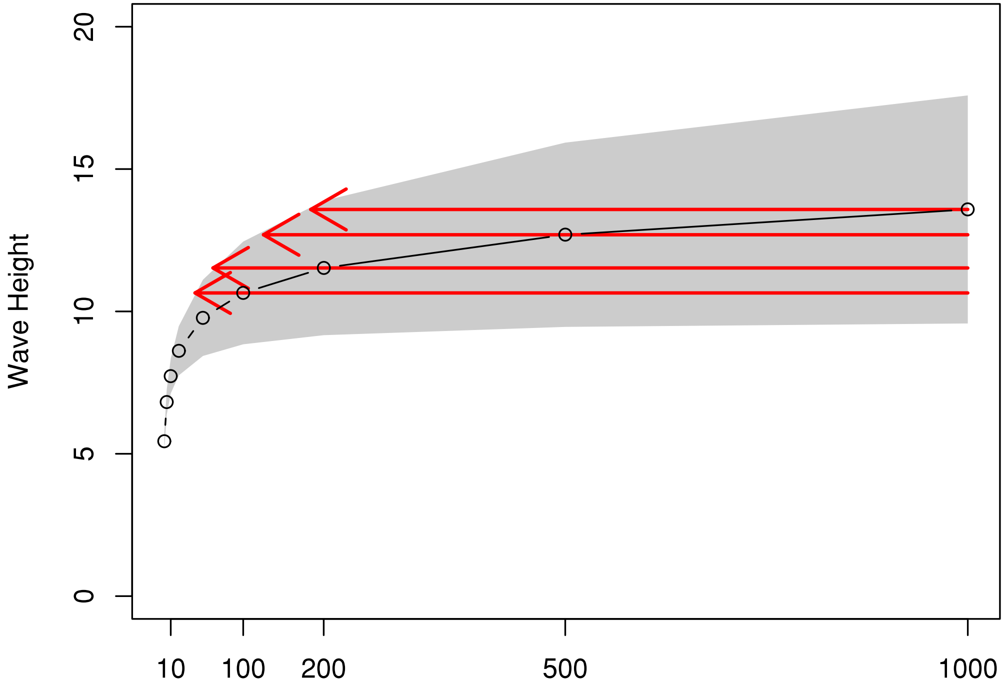



だって，１００年も１０００年も，信頼区間の幅の
大きさに大差が無いのだもの，　．．．

結局のところ，信頼区間を用いると，相対的な検討しかできません(どの位大きくなるとマズイの？)．

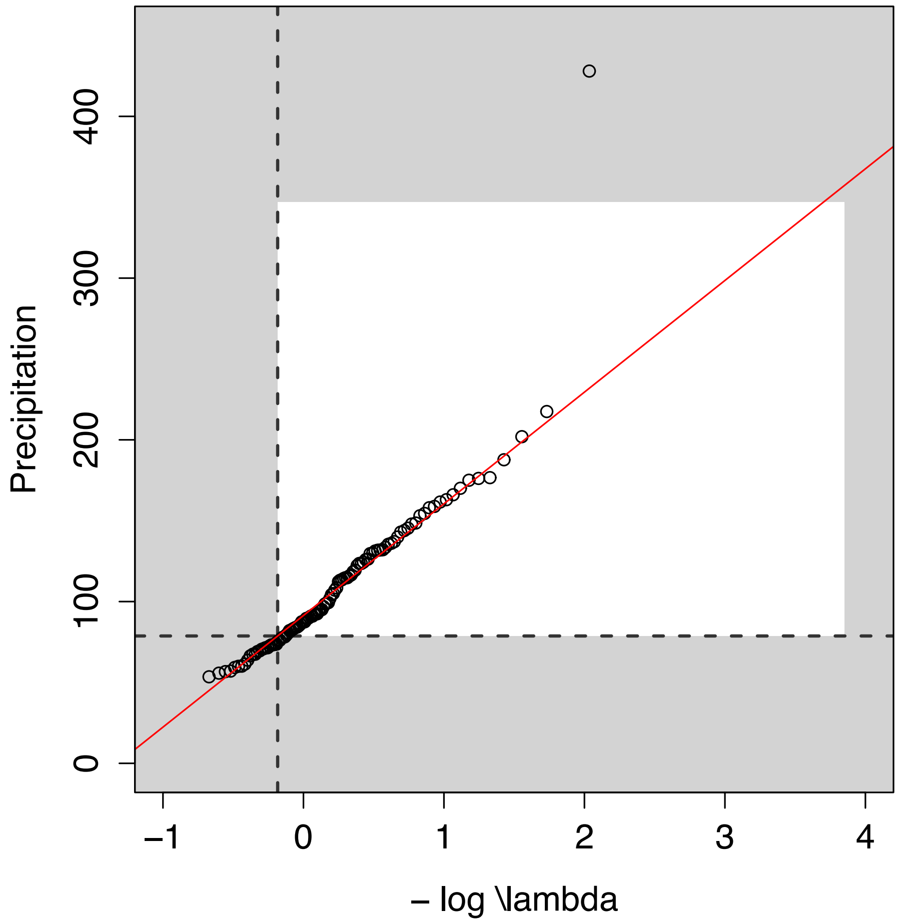


# タテの幅ではなくて，ヨコの幅に注目すると，　…

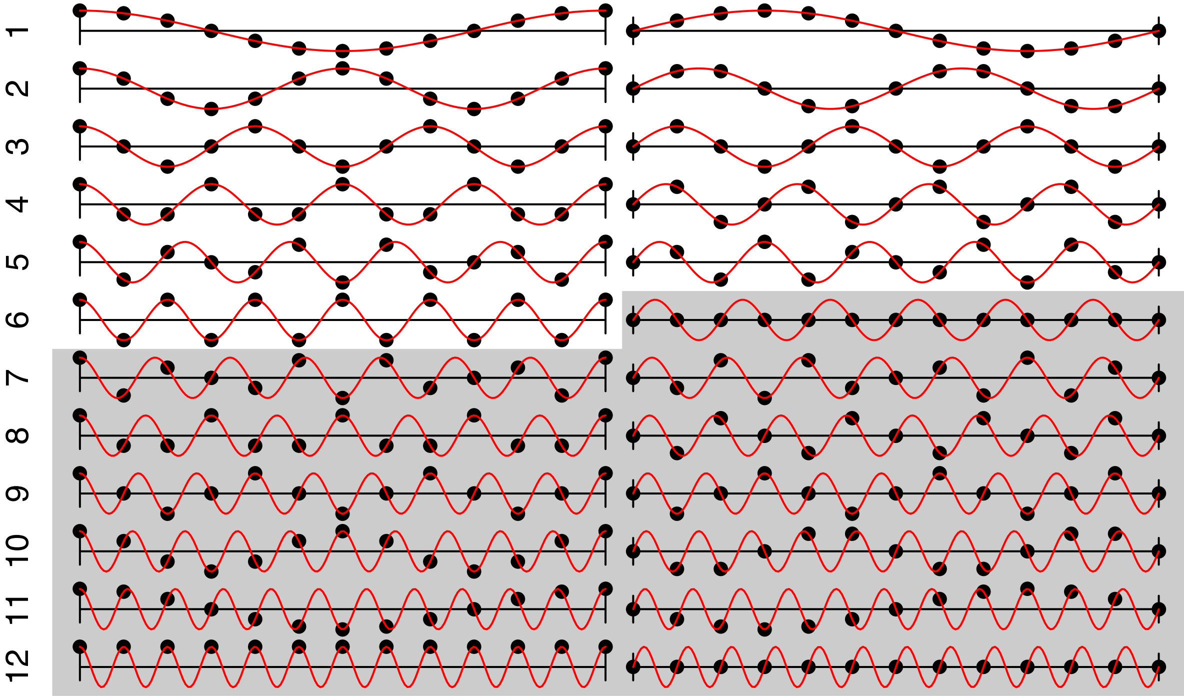



# 幅の片側が見えない位に広い（とてつもなく）

結論として，「容枠を設けよう」というのが，我々の提案である．

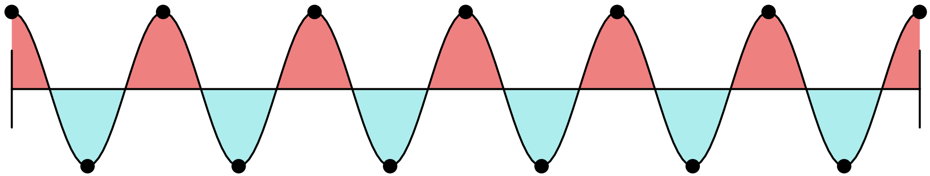


# Rule of Two

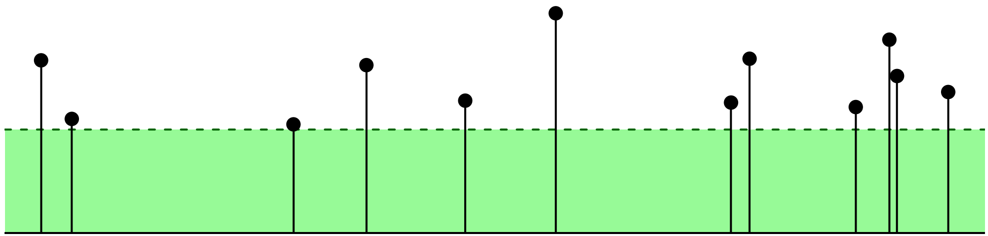



Nyquist Rule in Wave Measurments

# Rule of Two



Wave:  Crest & Trough

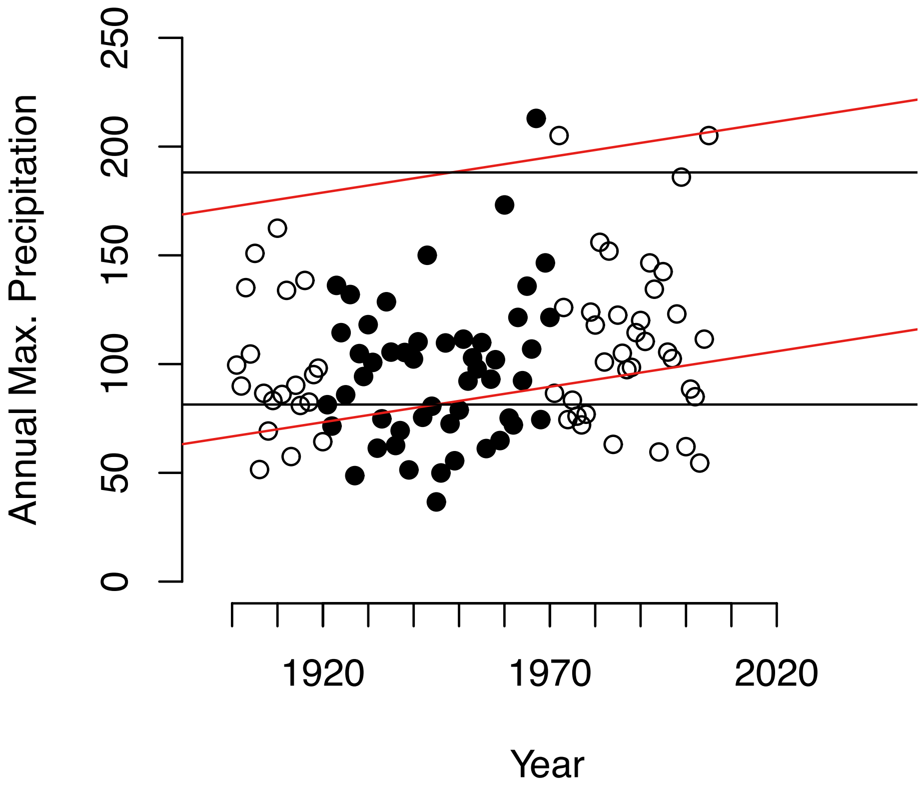


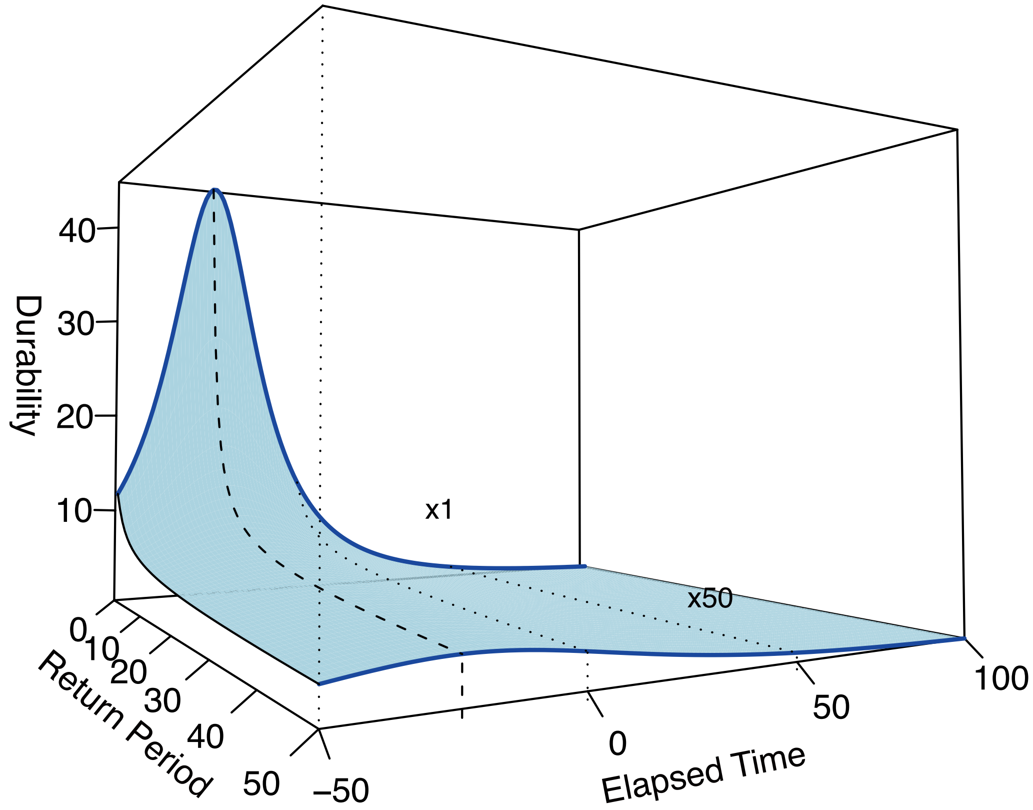
Poisson Events:  Start & End

# 2つの外挿限界

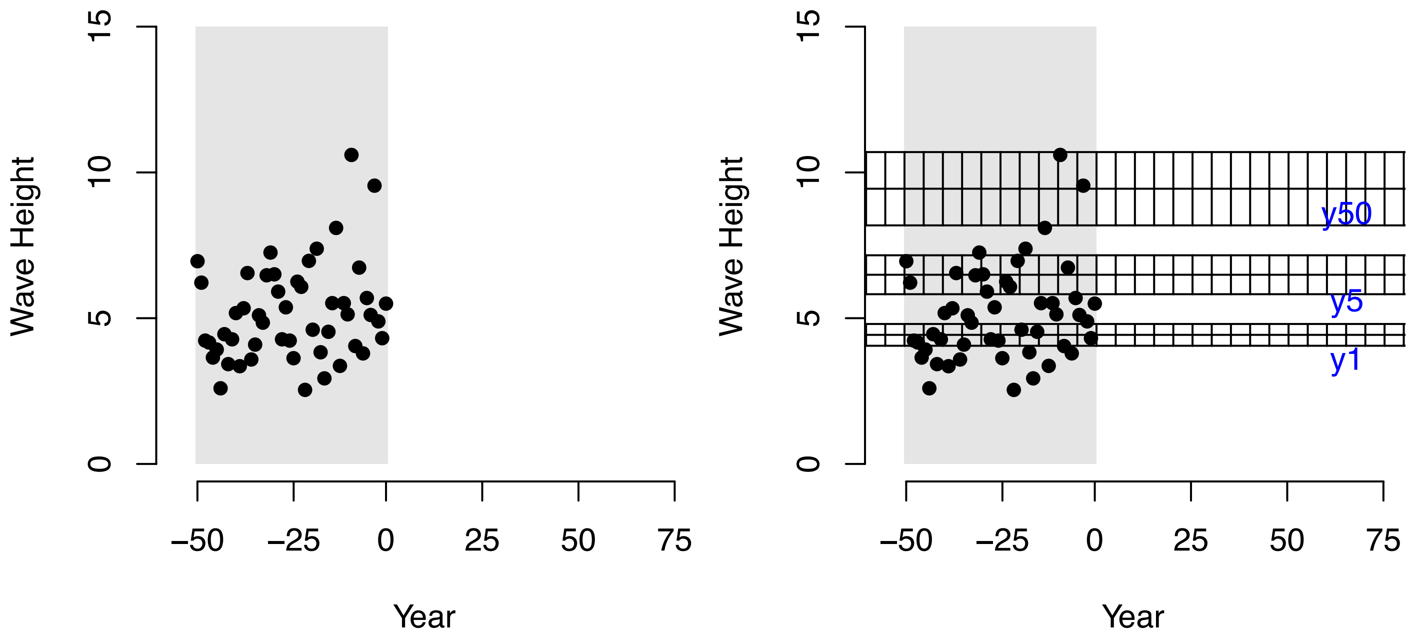
- 再現期間の長い希少確率の限界
- 経過時間に伴うモデルの適用限界

例：県の降雨量データ



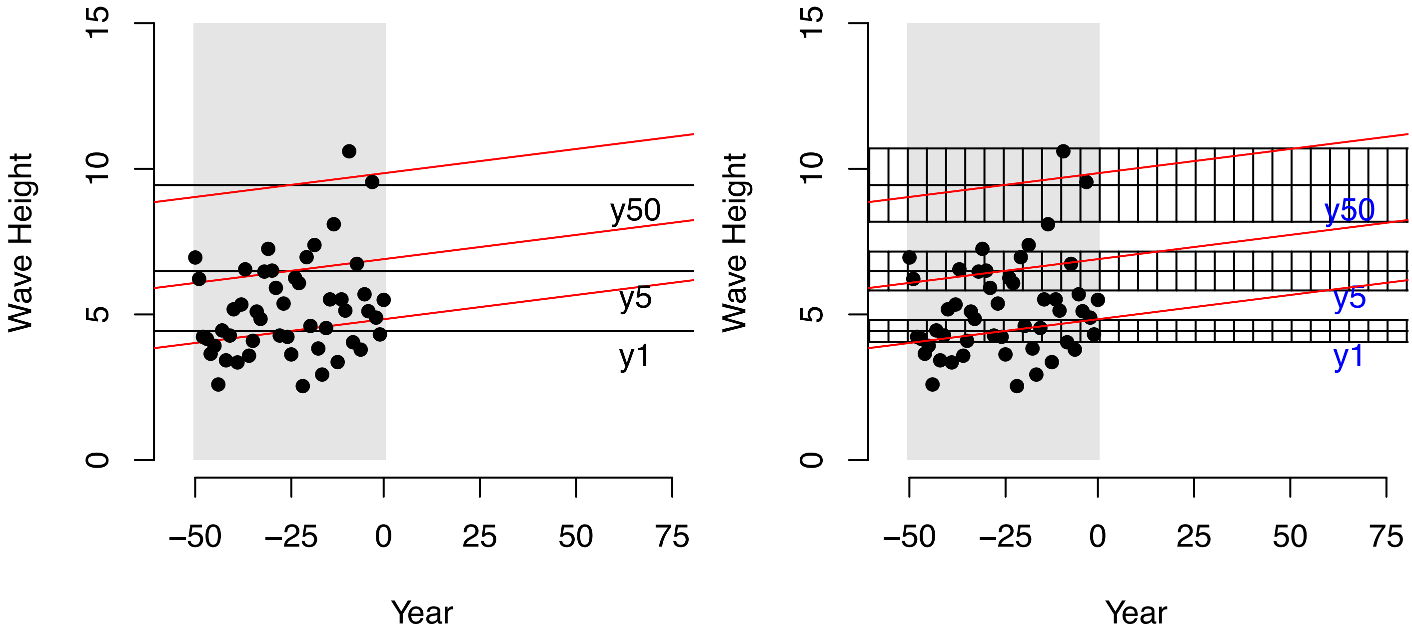
We can show a time history for extremes, and the estimated return levels with CI in conventional way. But we believe it?



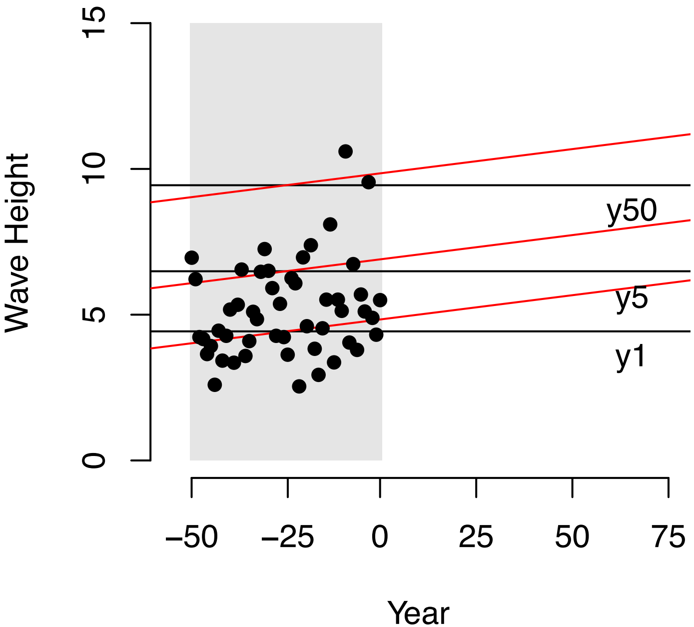
This is NOT tolerant of the probable trend.

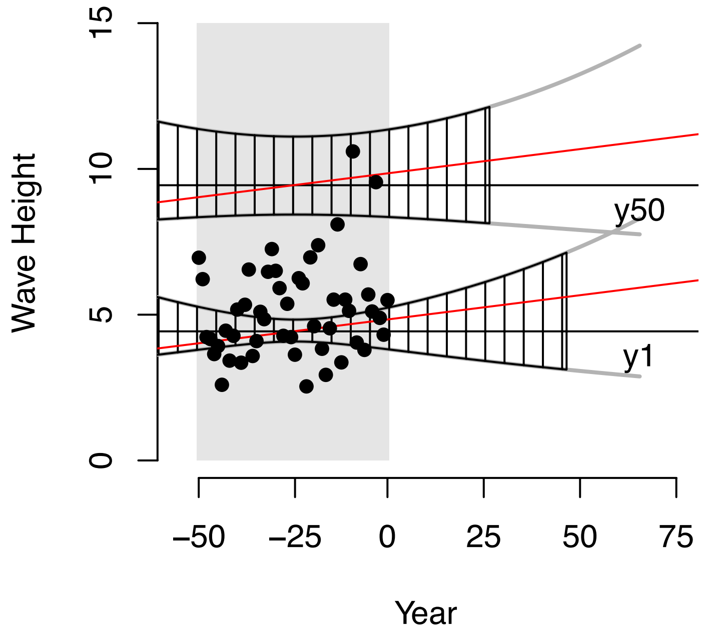
Stationary CI is very weak for the probable trend.

A reviewer also pointed out ... the *peculiar* properies ...



Our solution is
stationary estimation with non-stationary CI.
Thus, It is tolerable for the probable trend.



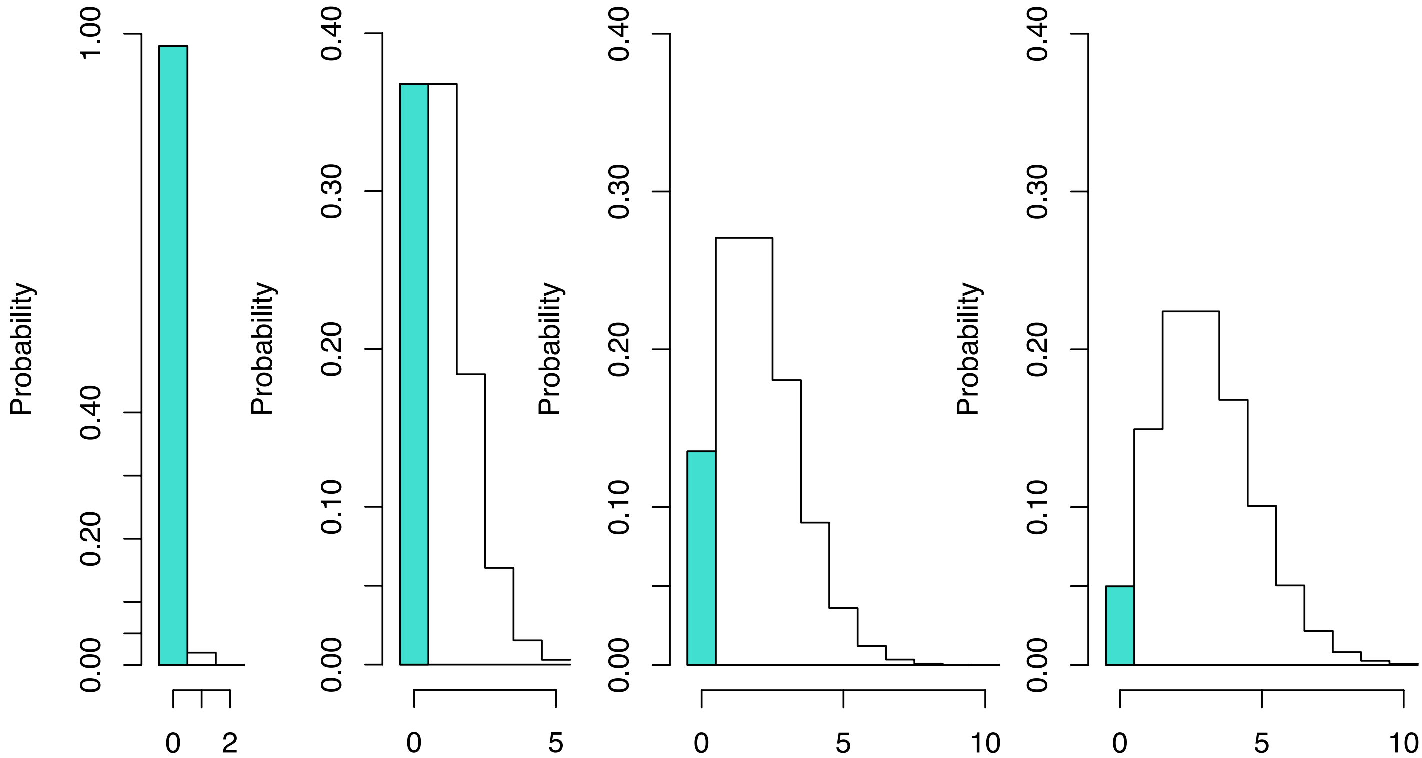
# 極値統計学を理解するための用語集（案）

回帰と共変量（Covariate）
緩慢変動関数と正則変動関数
ガンマ分布（ポアソン分布の共役分布として）
区間推定と信頼係数
グンベル分布とガンベル分布，一般（化）極値分布
経験度と（正規標本に対する）自由度
再現レベル（確率外力）とクォンタイル
再現レベルの誤差分散（＝尺度母数の自乗／経験度）
母平均の推定誤差分散（＝母分散／標本サイズ）
再現期間（この逆数を超過確率，生起率？）
再生性（確率変数の和の分布の）
残差，誤差，標準誤差，標準偏差
指数分布族と十分統計量
順位統計量と順位相関係数
順序統計量と上位 r 番目までの極値分布
情報行列（Fisher 情報行列）
周波数と周期の関係（生起率と再現期間の関係）
新記録と極値（待ち時間の特性の違い）
正則条件（MLE の）
漸近収束（極値分布への）
漸近正規性（推定量の）
漸近不偏性，不偏性，一致性
遭遇確率（単年あたりの超過確率とは異なる）
対数ピアソンIII型分布（ピアソン頻度分布の関数族）
大数の法則（弱法則と強法則）
ティ分布，カイ自乗分布，エフ分布（統計量の標本分布）

度と回（２度としないで！と５回もやっちゃった？）
内挿と外挿
中心極限定理
フレシェ分布
プロッティングポジション
母数（パラメータ）と統計量
母分布と標本
無情報事前分布（ベイズ推定）
有意水準と p 値
予測と推測
モーメント（積率モーメントとＬモーメント）
リンク関数（GLM（一般化線形モデル）における）
ワイブル分布と（逆）ワイブル分布

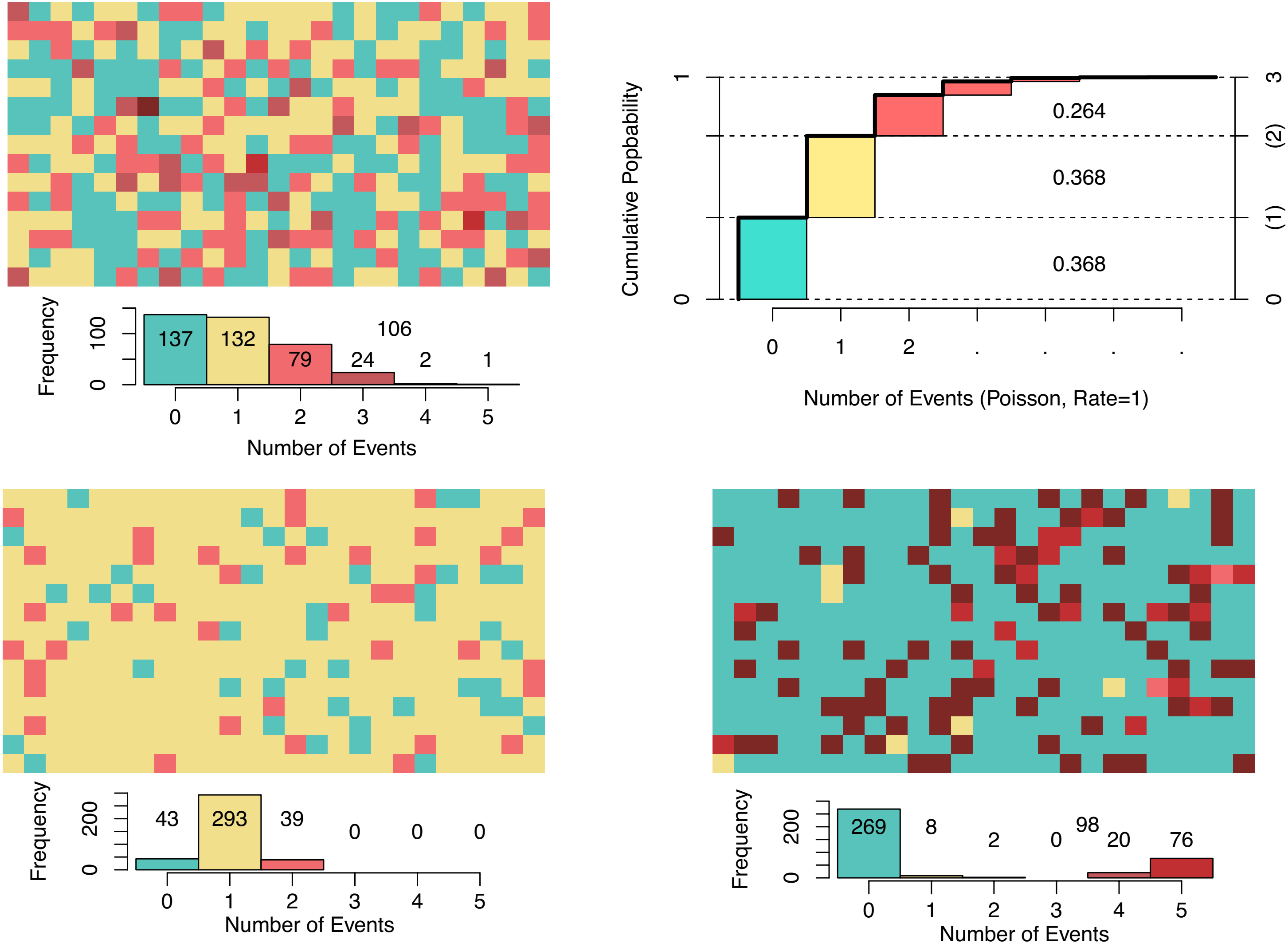
Block Maxima（Annual Maxima など）
Copula（接合分布関数）
Cramer-Rao の不等式
Extreme value index と Extremal index
Intensity measure と生起率（Occurrence rate）
Mean, Mode, Median（平均値，最頻値，中央値）
Mean excess と Mean residual life
Max stable（最大値安定性）と non degenerate（非退化）
POT (Peaks Over Threshold)
Profile likelihood
PP プロットと QQ プロット
Rule of three と Rule of thirds

# Rule of Three




$\lambda = 1/50$ $\lambda = 1$ $\lambda = 2$ $\lambda = 3$

# Rule of Thirds

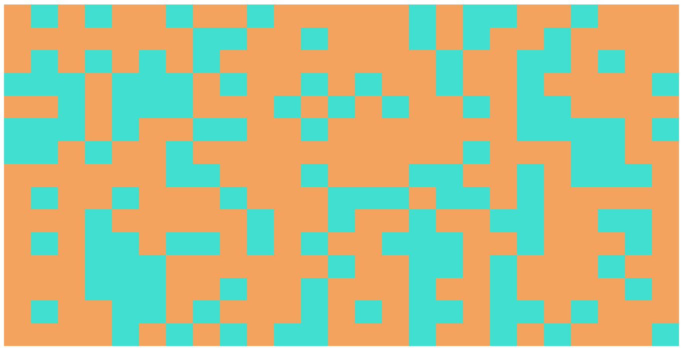


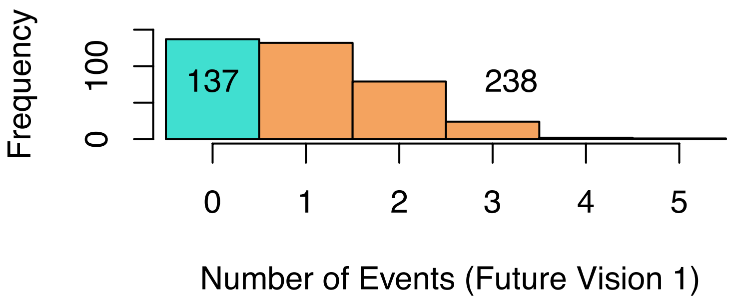
Frequency
0
100
137
132
79
24
2
1
106
0
1
2
3
4
5
Number of Events
Cumulative Popbability
0
1
0.368
0.368
0.264
0
(1)
(2)
3
0
1
2
. . . .
Number of Events (Poisson, Rate=1)
Frequency
0
200
43
293
39
0
0
0
0
1
2
3
4
5
Number of Events
Frequency
0
200
269
8
2
0
98
20
76
0
1
2
3
4
5
Number of Events

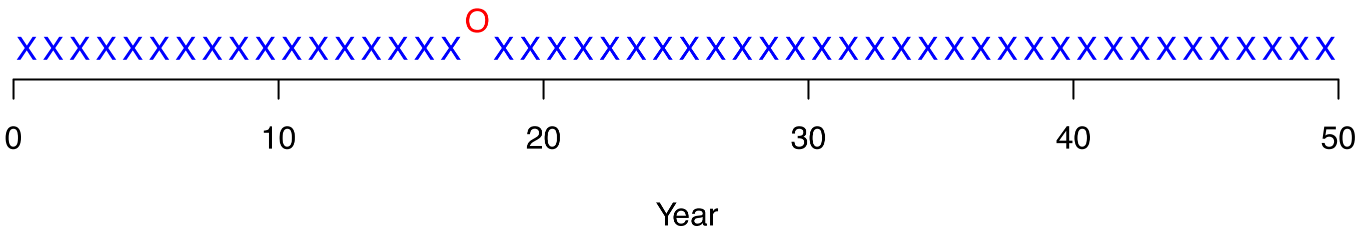
遭遇確率： $$Q_L = 1 - \left(1 - \frac{1}{R}\right)^L \approx 1 - \exp\left(-\frac{L}{R}\right)$$

$$Q_L\Big|_{L=R} \approx 1 - \frac{1}{e} = 0.632$$

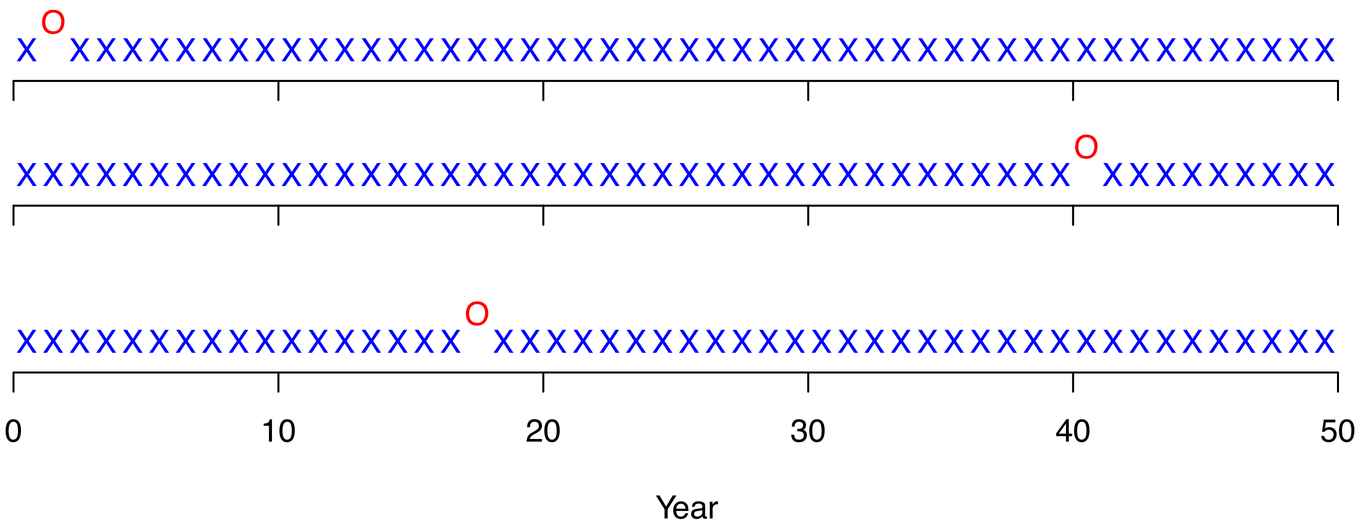
Future Vision 1

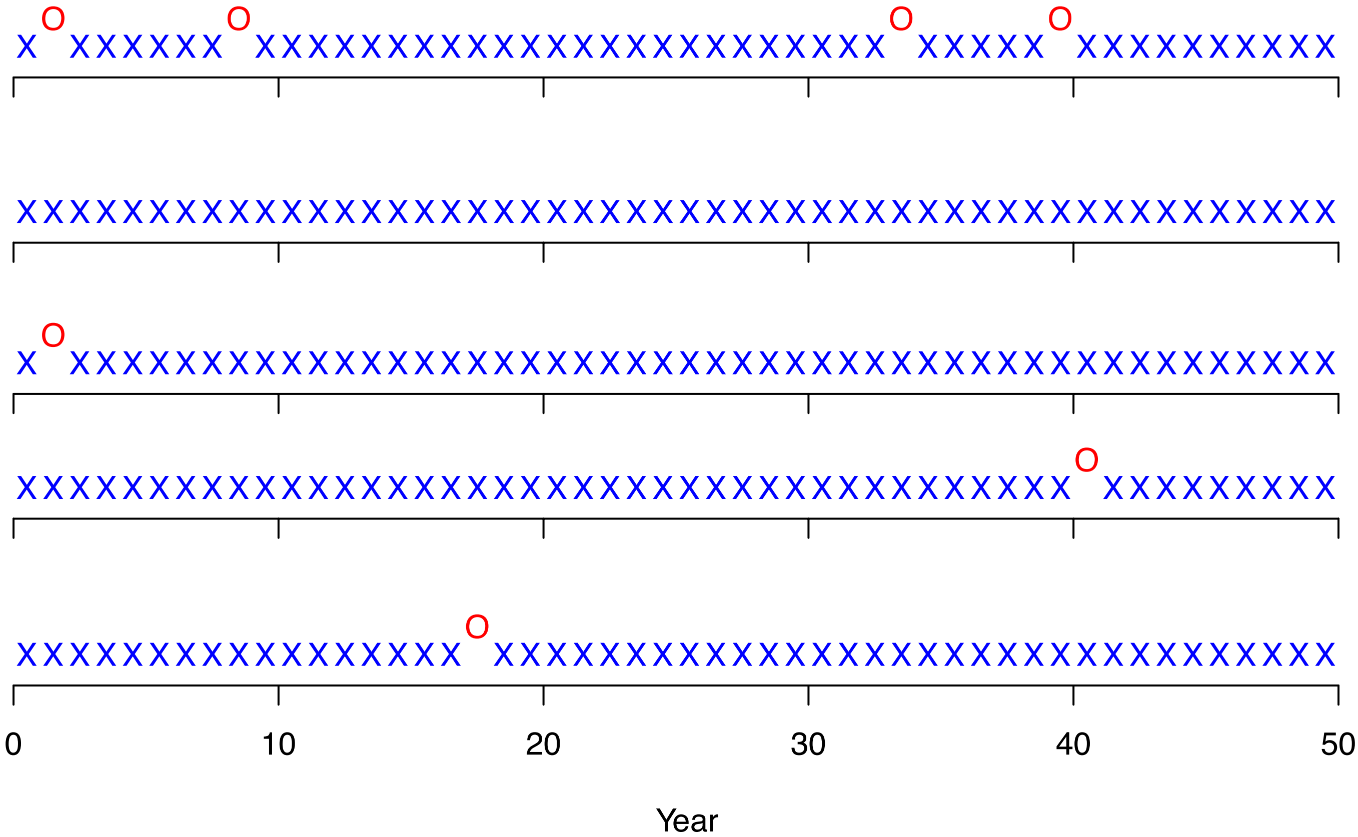






0
10
20
30
40
50
Year

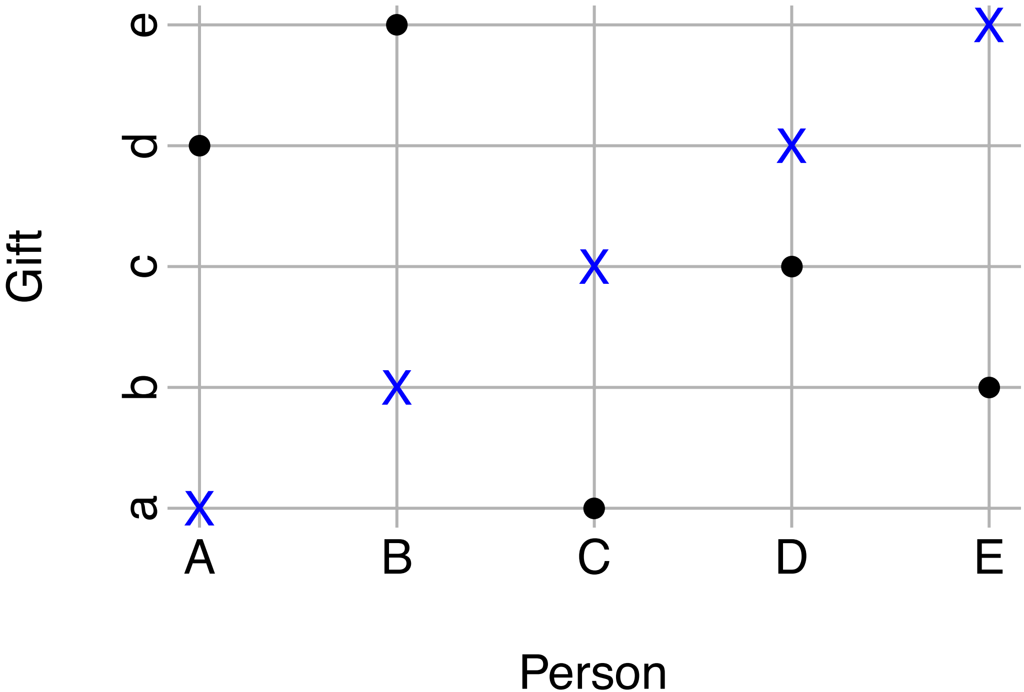

0
10
20
30
40
50
Year



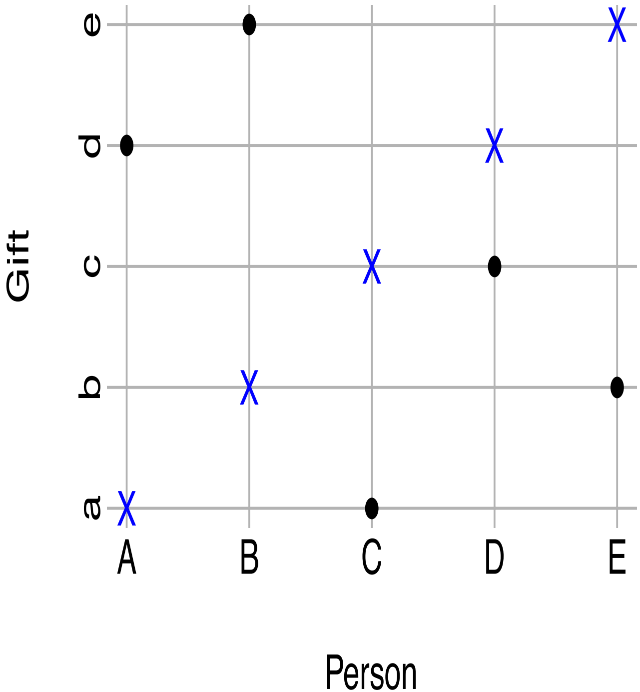

(Coin Toss) モデルが単純すぎて，確率を実感できない，あるいは，
1/50 という確率が作為的に導入ゆえ，「あくまでも理屈」の話にきこえる．

# Tragedy in Xmas Party

# Tragedy in Xmas Party



# Tragedy in Xmas Party




= Problème des Rencontres

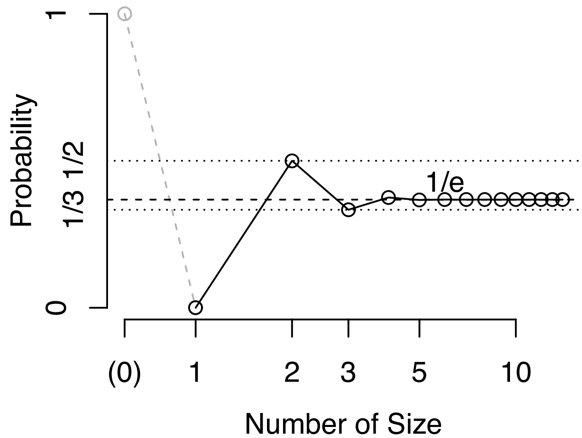

$$P_n = \frac{\Delta^n 0!}{n!}$$

| | 0! | 1! | 2! | 3! | 4! | 5! | 6! | 7! |
|---|---|---|---|---|---|---|---|---|
| D0 | 1 | 1 | 2 | 6 | 24 | 120 | 720 | 5040 |
| D1 | 0 | 1 | 4 | 18 | 96 | 600 | 4320 | |

$$P_n = \frac{\Delta^n 0!}{n!}$$

| | 0! | 1! | 2! | 3! | 4! | 5! | 6! | 7! |
|---|---|---|---|---|---|---|---|---|
| D0 | 1 | 1 | 2 | 6 | 24 | 120 | 720 | 5040 |
| D1 | 0 | 1 | 4 | 18 | 96 | 600 | 4320 | |
| D2 | 1 | 3 | 14 | 78 | 504 | 3720 | | |
| D3 | 2 | 11 | 64 | 426 | 3216 | | | |
| D4 | 9 | 53 | 362 | 2790 | | | | |
| D5 | 44 | 309 | 2428 | | | | | |
| D6 | 265 | 2119 | | | | | | |
| D7 | 1854 | | | | | | | |

$$P_n = \frac{\Delta^n 0!}{n!}$$

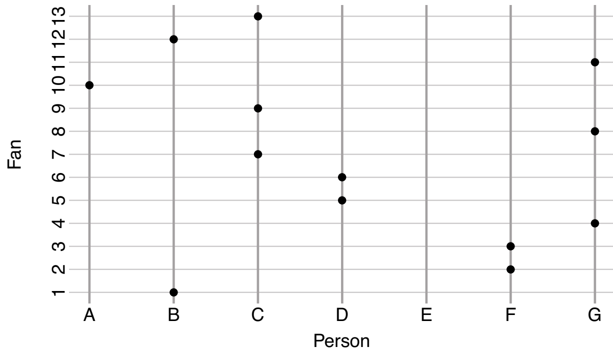



| n=0 | n=1 | n=2 | n=3 | n=4 | n=5 | n=6 | n=7 |
|---|---|---|---|---|---|---|---|
| 1.000 | 0.000 | 0.500 | 0.333 | 0.375 | 0.367 | 0.368 | 0.368 |

exp(-1) = 0.3678794...

# Fan Voting for Idols

# アイドルファン投票の悲劇




ファン投票の例示（アイドル E に残念な思いが）

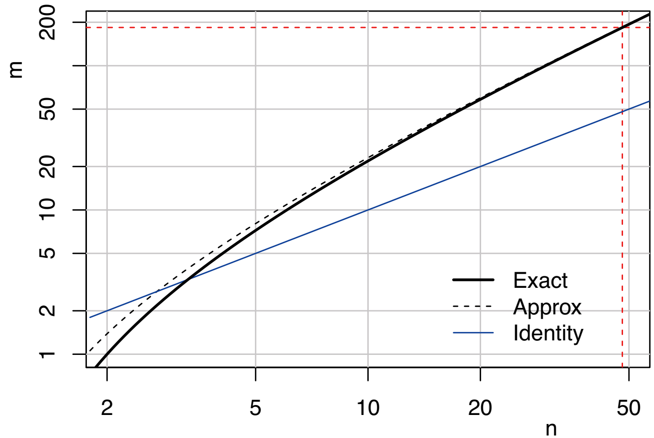
$g(n) = n^m$ とすると，$Q_{(n,m)} = \dfrac{\Delta^n 0^m}{n^m}$,　$S_j = \dbinom{n}{j} \dfrac{(n-j)^m}{n^m}$ となる.

$n^m$ は，区別のできる $n$ 個の壷に区別のできる $m$ 個の玉を入れる総数である．$n = 48$ 人のアイドルにファン投票（$m$ = ファンの数）を行う際に，$Q_{(n,m)}$ は，ファンの一人からも投票されないアイドルがゼロである確率を表している．この場合，$m$ の数を変えれば，その期待値 $\lambda$ も変わる．あるアイドル 1 人に着目して，そのアイドルがファンの 1 人からも投票されない確率は $(1-1/n)^m$ であり，少なくとも 1 人のファンが投票してくれる確率は $1-(1-1/n)^m$ となる．そして，アイドル $n$ 人に対する期待値 $\lambda$ は，次式で与えられる．

$$\lambda = n\left(1 - \frac{1}{n}\right)^m$$

ファンの一人からも投票されないアイドルの数の期待値は（アイドル $n$ 人に対して，ファンの数 $m$），

$$\lambda = n\left(1 - \frac{1}{n}\right)^m$$



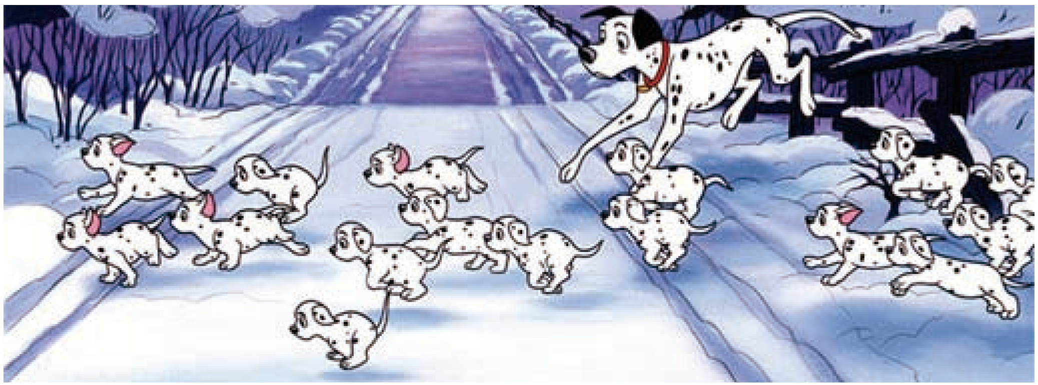

$\lambda \fallingdotseq 1$ となる n と m の組合せ

期待値 $\lambda = 1$ として，$n$ と $m$ の関係を示したものである．図中の点線は近似式 $m \approx n \log n$ と（$\lambda = 1$，$n \gg 1$）である．$n = 48$ の場合，$m = 184$ で，$\lambda = 0.997$ となる（残念ながら，この場合には，期待値をキッカリ 1 にすることは難しい）．

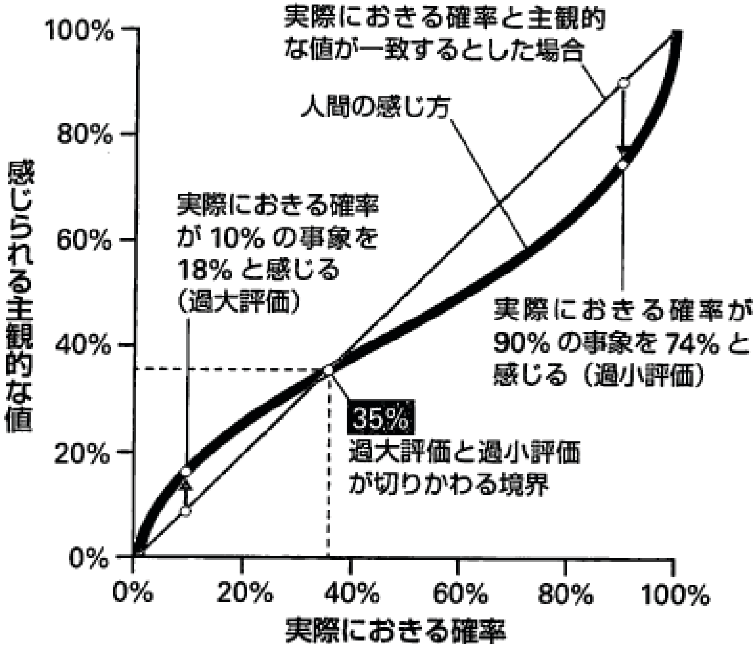
$Q_{(n,m)}$ を求めると，

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (n, m) = | (5, 7) | (7, 13) | (13, 32) | (48, 184) | (101, 463) | (365, 2150) | (2013, 15310) |
| Q(n, m) = | 0.2150 | 0.2973 | 0.3091 | 0.3494 | 0.3544 | 0.3639 | 0.3671 |
| lambda = | 1.0486 | 0.9436 | 1.0036 | 0.9974 | 1.0082 | 1.0014 | 0.9999 |

となり，$1/e \approx 0.368$ にゆっくりと収束する様子がわかる．この場合は，$n$ にかなり大きな数が必要であり，$n = 48$ でもまだ十分とはいえない．しかし，数が大きくなるにつれて，計算の労が急激に増える（差分で計算できるのも $n = 13$ 程度までであろう）某女子グループの数に合わせたのは，そのような数理的な背景がある．なお，実際の“総選挙”では，投票されるアイドルの数も 48 より多いが，桁違いの投票数（「私のことは嫌いでも、AKB48 のことは嫌いにならないでください」の名言で知られる 2011 年には $n = 150$，総投票数 $m = 1,166,145$）ゆえに，事実上 $\lambda \approx 0$（$\approx 0.343/10^{3385}$）となる．その分母は，無量大数 $10^{68}$ や googol $10^{100}$ よりも大きい（余談であるが，大きな数のことを**天文学的数**と言われることが多いが，**組合せ論的数**とも表現してもよいのかもしれない．なお，不可説不可説転という超巨大数もある）．



# 物事が起きたときの感じ方の大小




$$0.37：0.63 \fallingdotseq \mathit{0.35}：\mathit{0.40}\ \ ???$$

（Newton, 2013年6月号）

# いろんなところに現れる
# 約 2/3 という割合

海：陸＝7：3

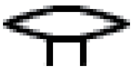
＝360:150

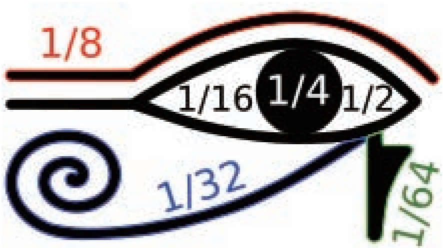
(×$10^6$ $km^2$
メガ平方キロと読む？)

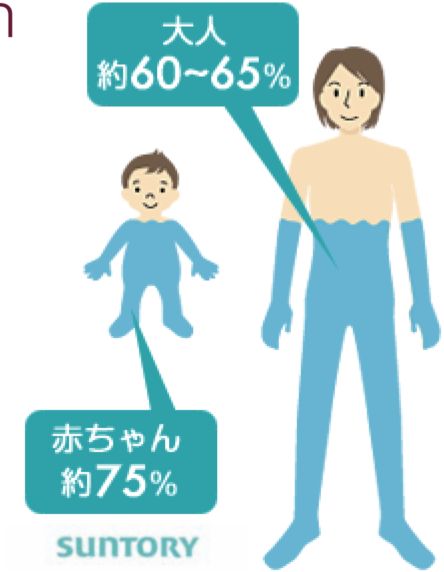
## Egyptian fraction

ex. $\frac{2}{5} = \frac{1}{3} + \frac{1}{15}$

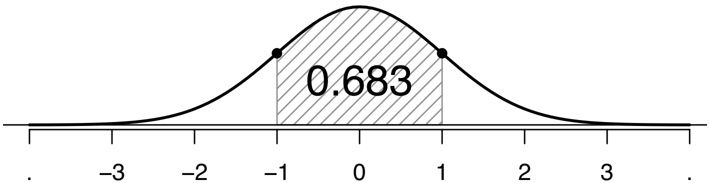
$\frac{1}{7} = \frac{2}{3} + \frac{1}{30}$

 $= \frac{2}{3}$




正規分布ですら，　…




2014年(平成26年)10月17日　金曜日　朝日新聞

# 火山列島に生きる

土岐憲三さん

諦観とは、自然災害を「天災」という言葉で呼ぶことにも表れているように、天から来るものだからしゃあないし、めったに起きるものでもない、そんな風に考えることです。

運が悪かった、ではなく、どうすれば死なずにすんだか、徹底的に考えて実践することが大切なのです。先日の朝日新聞が、地震の揺れを検知して電気を遮断し、電気火災を防ぐブレーカーの普及が進まないと報じ

また、私が懸念するのは、科学が楽観を強めかねない、ということです。私の専門の地震災害については近年、地震発生の切迫度を確率で表すようになってきました。ところが海溝型巨大地震と比べて内陸型地震はるかに頻度が低く、発生確率は数%程度の小さい数字になりがちです。発生する可能性はこんなに小さいんだと、安心させる情報になりかねません。

```
> 1 - (1 - 8/100)^9
[1] 0.5278386
```

阪神大震災後、発生確率を事前に予測したらどうなったか計算したら、最大限に見積もっても8%でした。事前にこの数字を発表しても、多くの人は、「大丈夫じゃないか」と逆に安心したのではないでしょうか。

御嶽山の今回の噴火についても、噴火警戒レベル1という数字が安心情報となっていたのかもしれません。科学的知見を、科学そのものの不確実性や限界も含めて、社会にどう伝えるか。それは大きな課題です。

自然をそのまま受けいれる。日本文化のよいところではありますが、やはり、自分たちが向かい合う自然に果たしてどれだけのリスクが存在するものか、個人も行政も明確に意識するべきです。

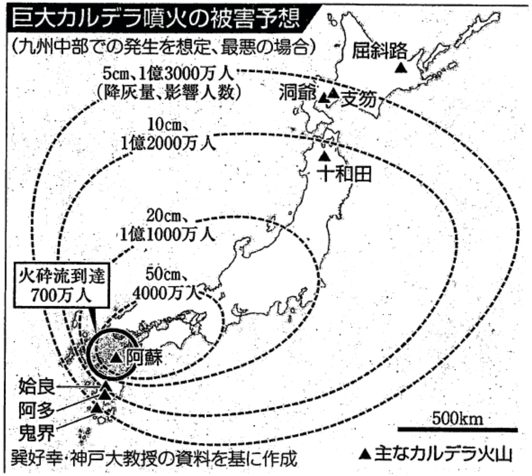
（聞き手・辻篤子）

土岐憲三さん

## 今後100年の発生確率試算

# 巨大カルデラ噴火1%

巨大なカルデラ（陥没地形）をつくる巨大噴火が今後100年間に日本列島で起きる確率は約１%とする試算を神戸大の巽好幸教授（マグマ学）らがまとめ、22日発表した。最悪の場合、１億2000万人が死亡し、実質的な「日本喪失」もありうるとしている。

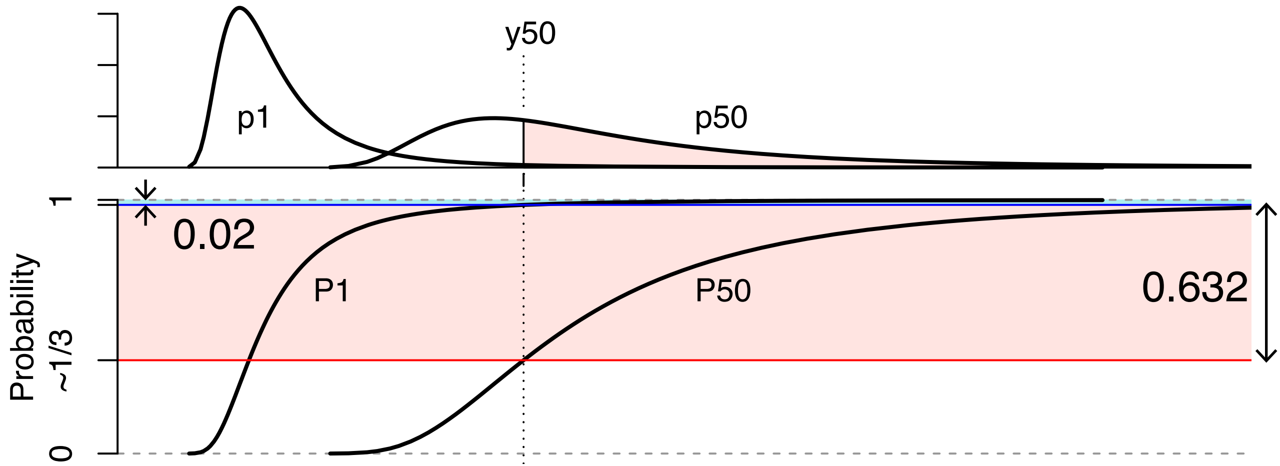



百年で1%という値について巽教授は「決して低い数字ではない。いつ起きてもおかしくない。覚悟が必要だ」と話し、深さ数キロ程度のマグマの様子を観測できる技術を緊急に開発することを訴えている。研究成

```
> 1 -  (1 - 0.01/100)^7000
[1] 0.503
```

ののちゃん
作・いしいひさいち
6014


(平成26年)6月21日



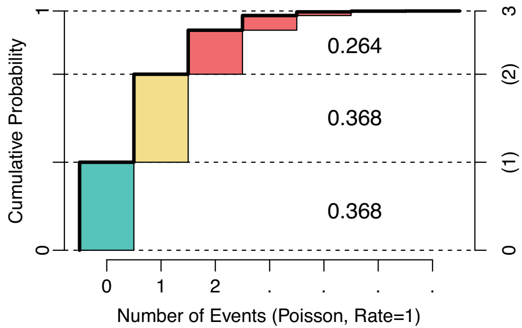

# 甚大災害が生じる確率は希少か？ ~ 防災の日に思う

「**科学の常識は，確率外力の非常識なのか，確率外力の常識が，科学の非常識なのか？**」なんともビックリさせるような言い方をしているが，説明不足ゆえに意味不明であろう．

たとえば，近年 30 年間の年最低気温の平均が，50 年前のそれとの差がゼロである仮定のもとで，現実に得られた**データを用いて求められた差がゼロからどれだけ乖離しているかを確率で計る**．すなわち，データから得られた差以上に大きな値をとる確率が 0.05 よりも小さければ，**先に定めた仮定のもとでは，0.05 というわずかな確率でしか生じ得ないことが，現実に起こったことになる**．しかし，**そのような希少確率で生じると考えるよりも，むしろ，仮定がおかしいと考えて**，その仮定が正しいとする仮説を棄却し，近年の年最低気温の平均は昔とは異なる，と判断するのが現実的である．少しメンドクサイ論理であるが，結果的には棄却する仮説を予め用意するのである．これが統計学のロジックであり，**現実世界で起こりうる偶然としては，ありえない希少確率を 0.05 ととるのが**，科学の常識である．

これに対して，豪雨や高潮，高波などの風水害の原因となる**自然外力の来襲確率に，0.02 や 0.01 という値**が用いられる．すなわち，50 年に平均 1 回の頻度で来襲する外力が 1 年間に生起する確率は 1/50 として扱い，その来襲外力の規模，すなわち，豪雨を対象にしているなら日降雨量，高潮なら潮位偏差を求める．これを確率外力とよび，外力の防御を目的に整備される堤防や防波堤の供用期間を 50 年もしくは 100 年として，**0.02 や 0.01 の確率で来襲する外力を検討するのが一般的**である．1000 年の津波の再来に備える際には，0.001 の確率を考えていることになる．

賢明な読者は既にお気づきであろう．「**科学の常識であれば棄却されうる希少確率 0.02** の頻度で来襲する外力に対して，**堤防や防波堤を整備するのは非常識か？**」というのが，冒頭の問題提起である．もちろん，その疑問文を否定するのが正しい．しかし，ひょっとすると誤解が生じてもおかしくないような話である．　甚大な被害を及ぼす自然外力を 1 年単位で数えることが誤解の原因になっている．むしろ，**50 年に平均 1 回来襲する外力は，50 年単位で考えるベキである**．そうすれば，そのような規模の外力が来襲しない確率は約３分の１，ちょうど１回来襲する確率も約３分の１で，２回以上来襲する確率も約３分の１となる．もはや，わずかな確率ではなく，**科学の常識と確率外力の常識が背反しない**．また，１回以上来襲する確率は，来襲しない確率の倍程度に大きく，その外力規模に備えた対策は，おおいに意義がある．




# 御用！

1）美咲さんと考える

2）差異を平均で見るとき
　にはｔ検定？

3）不確からしさの度合い