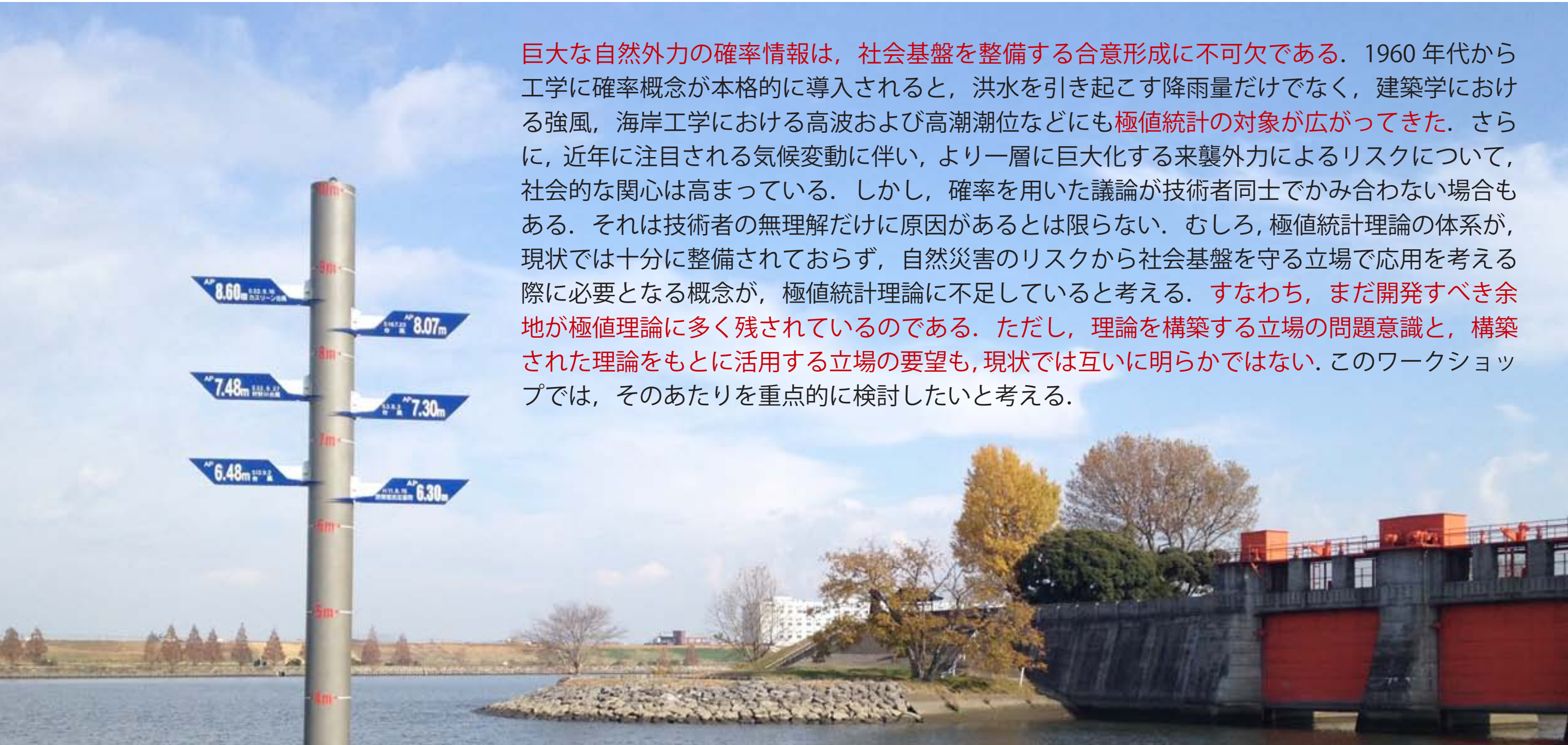
文部科学省委託事業 数学協働プログラム（受託機関：統計数理研究所）

甚大災害の外力想定に必要となる極値統計解析法の背景と活用

http://coop-math.ism.ac.jp/event/2014W09

2014 年 12 月　8 日（月）／京都大学 防災研究所

主催：**名古屋工業大学 高度防災工学センター，京都大学防災研究所，統計数理研究所**

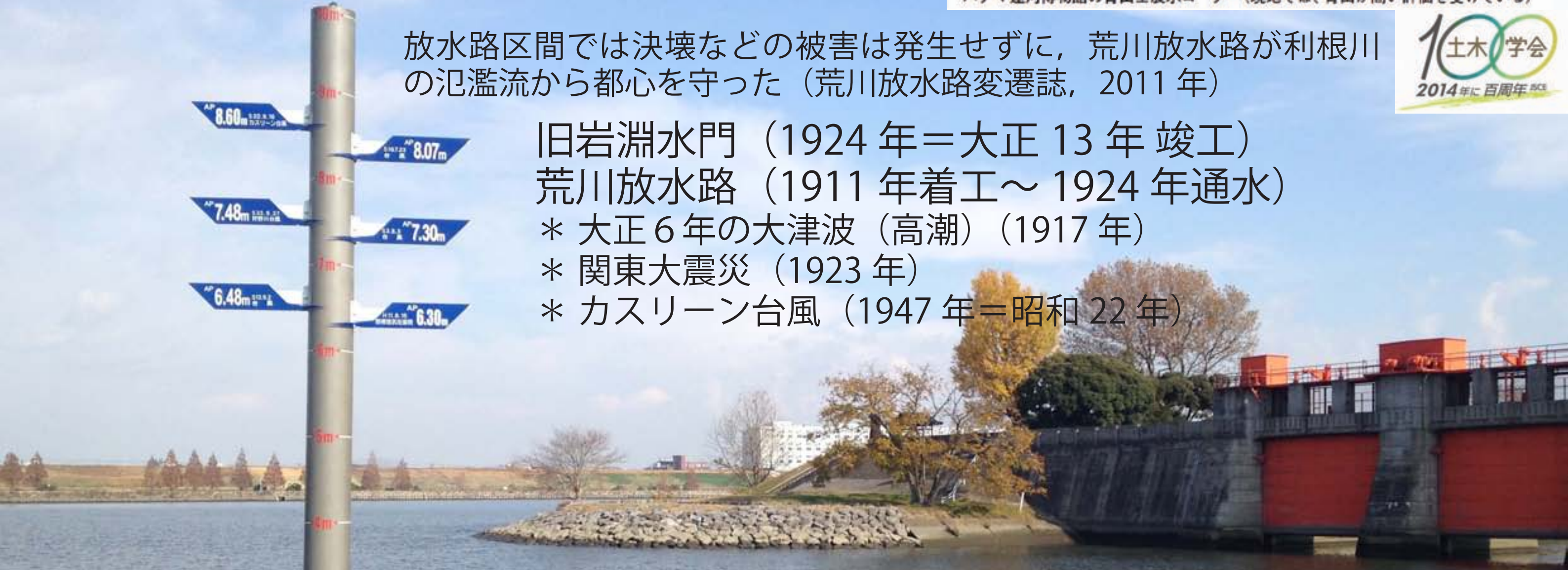


巨大な自然外力の確率情報は，社会基盤を整備する合意形成に不可欠である．1960 年代から工学に確率概念が本格的に導入されると，洪水を引き起こす降雨量だけでなく，建築学における強風，海岸工学における高波および高潮潮位などにも極値統計の対象が広がってきた．さらに，近年に注目される気候変動に伴い，より一層に巨大化する来襲外力によるリスクについて，社会的な関心は高まっている．しかし，確率を用いた議論が技術者同士でかみ合わない場合もある．それは技術者の無理解だけに原因があるとは限らない．むしろ, 極値統計理論の体系が, 現状では十分に整備されておらず，自然災害のリスクから社会基盤を守る立場で応用を考える際に必要となる概念が，極値統計理論に不足していると考える．すなわち，まだ開発すべき余地が極値理論に多く残されているのである．ただし，理論を構築する立場の問題意識と，構築された理論をもとに活用する立場の要望も, 現状では互いに明らかではない. このワークショップでは，そのあたりを重点的に検討したいと考える．

“ 出会った時よりも良い世界を未来へ渡すために ”
ジョン・ハーシェル卿

資料出典：土木学会誌，2008年8月

パナマ運河博物館の青山士展示コーナー（現地では、青山が高い評価を受けている）



放水路区間では決壊などの被害は発生せずに，荒川放水路が利根川の氾濫流から都心を守った（荒川放水路変遷誌，2011 年）

旧岩淵水門（1924 年＝大正 13 年 竣工）
荒川放水路（1911 年着工～ 1924 年通水）
＊ 大正６年の大津波（高潮）（1917 年）
＊ 関東大震災（1923 年）
＊ カスリーン台風（1947 年＝昭和 22 年）

文部科学省委託事業 数学協働プログラム（受託機関：統計数理研究所）

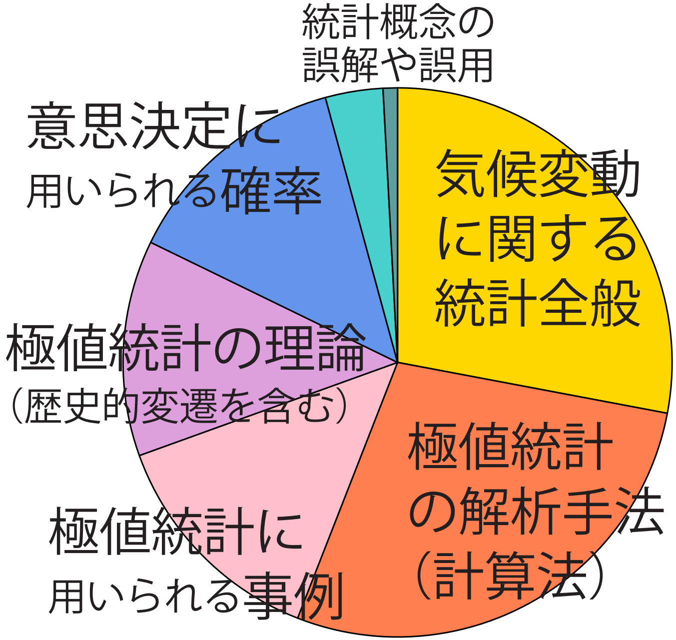
忘れられた科学 ～ 数学

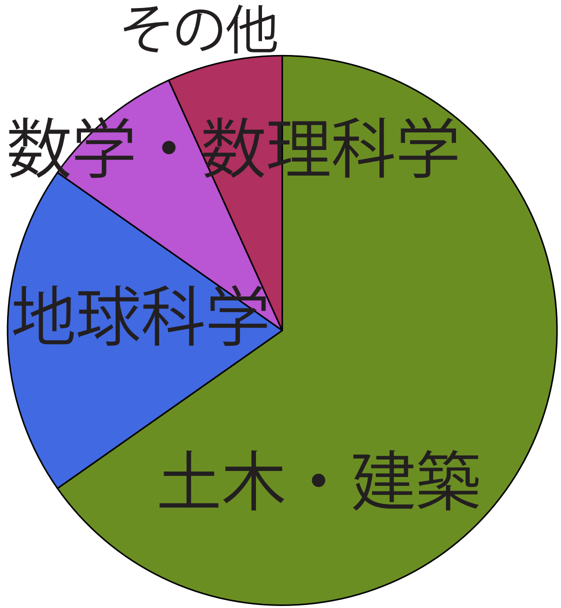
2006 年　日本学術会議シンポジウム「礎の学問：数学」

あなたの研究チームに数学をバックグラウンドに持つ人は含まれていますか？　はい２６％：いいえ７４％

あなたのライバルの欧米の研究チームに数学をバックグラウンドに持つ人は含まれていますか？

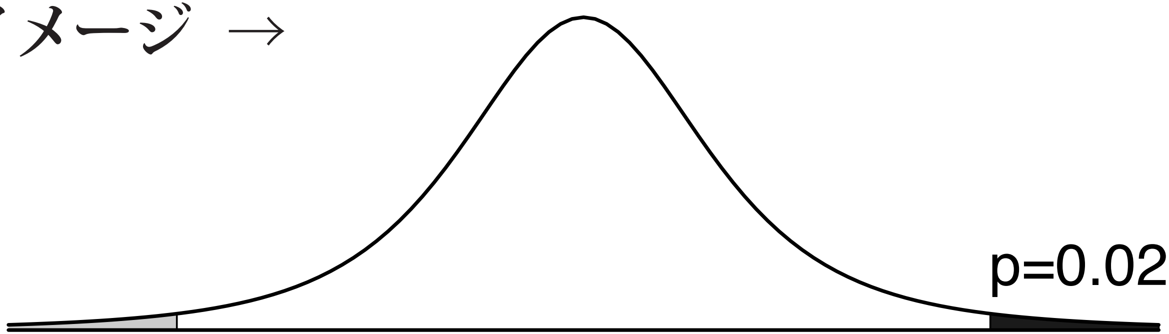
はい６２％：いいえ３８％



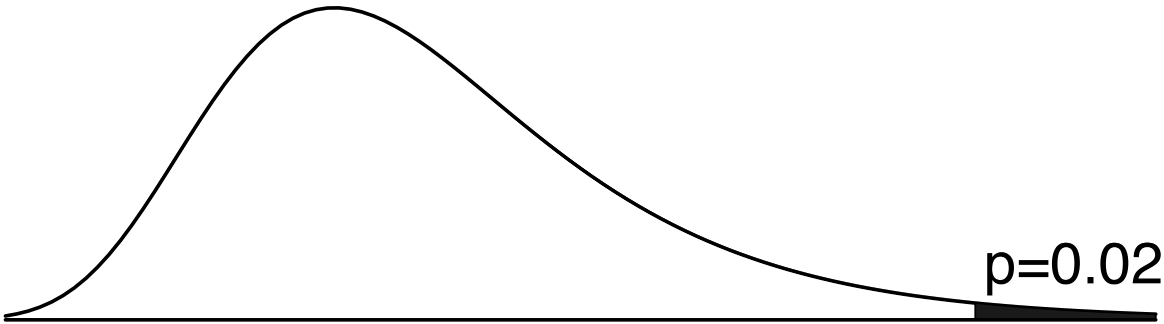


参加者の内訳 (12/6 時点の申込数 126 名)

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



EVA 2.0

～ 発展しつつある極値統計解析

（**E**xtreme **V**alue **A**nalysis）の次のステップに向けて

（趣旨説明）

名古屋工業大学
社会工学専攻／高度防災工学センター
統計数理研究所 客員
北野 利一

Gumbel 教授による金字塔となるテキスト “Statistics of Extremes” が出版された前後から，数理統計学において徐々に理論の精緻化がすすむとともに，応用上の計算手法も様々に開発される一方で，**応用分野である土木・建築の分野では独自の発展**をしてきた．そのため，**現在その乖離が大きい**ように感じる．**しかしながら，他方にだけ理があるというものではなく，どちらにも理があると考える．**極値統計解析に限らず，数理統計学では，提案された統計モデルでうまく行く場合を例示できるという立場にあり，うまく行かないモデルをあえて示すことは一般にしない．しかしながら，現実世界で要求される問題に対して，得られたデータをもとに問題解決を迫られる．その場合，極値統計理論や数理統計理論における知見を適用してもうまくいかないことが多いことも現実である．そのため，**あえて統計学を無視（？！）したような手段も講じられる**ことが多い．しかしながら，なぜそのような手段をあえて使ったのか？という理由が（数理的な論理で）明確にされることは少なく，**対象とした現実の問題解決の文脈と切り離してしまうと，第三者には分からなくなってしまう**．また，そのような**文脈依存である手段は，本来的には一般化できないと考えるが，独り歩きしている現実**もある．このような混沌とした状況を EVA 1.0 と呼ぶならば，発展しつつある極値統計解析の次のステップとして，

EVA 2.0 と呼ぶべき指針が必要である

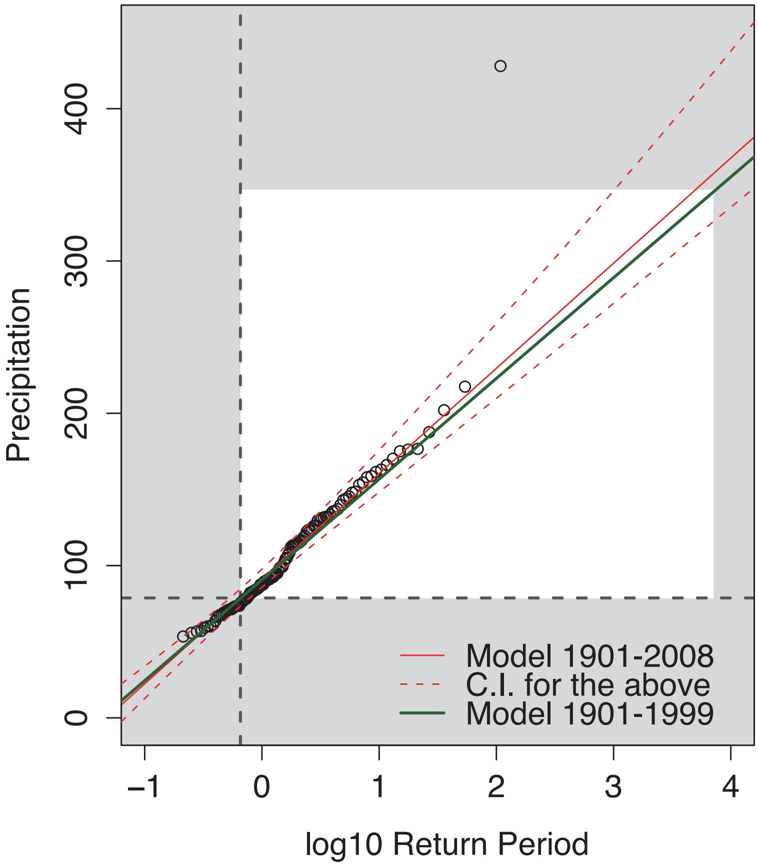
と考えている．**できる限り数理統計学に沿った形で，しかし，応用する分野からも満足できるようにするためには，どのように統合して考えて行けばよいか？** **今回のワークショップを起点に，近いうちに，その方向を示すことができればさいわいと考えている．**

極値データを解析する際の考え方の根底が異なるために，**議論が平行線になる場面が幾つかある**．そのような考え方の違いを具体的に示しておこう．**その中でも最も大きな相違点は，極値理論（極値に対する漸近理論）を重視するか否か**，すなわち，

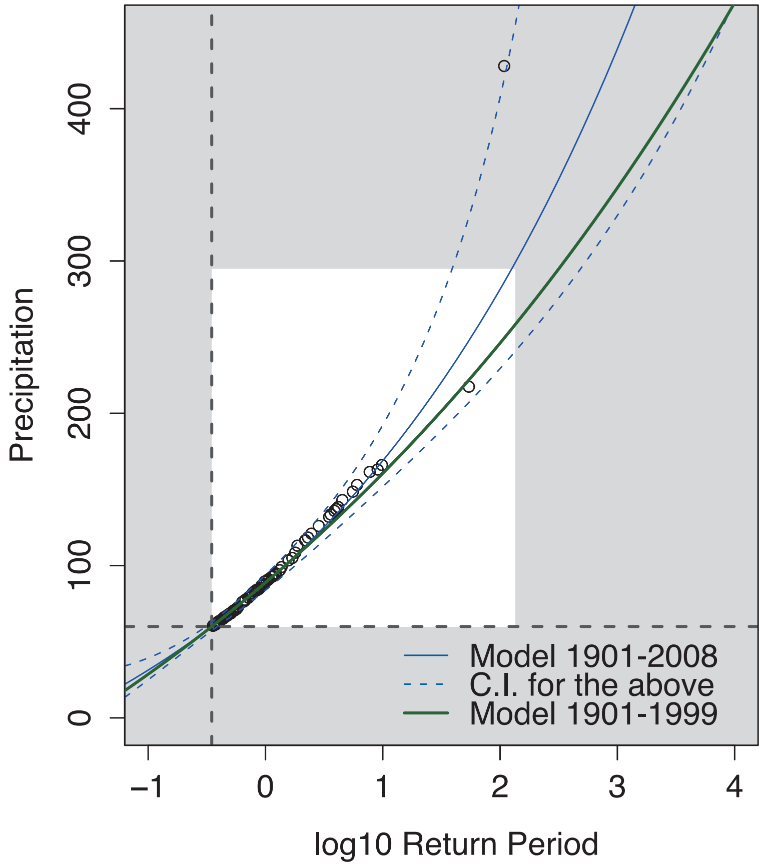
考え方の相違点１）極値統計解析を行う際，**漸近理論から得られる極値分布に限定して解析をすすめる立場**と，対数ピアソン III 型分布や（上限が無く無限大まで広がる）ワイブル分布などの極値分布では無い分布も母分布の候補として検討の対象に解析をすすめる立場の相違がある．小から大，すなわち，**検討する範囲を，理論分布となる小さな範囲で検討して，そこで不都合が生じるまで，検討の範囲を広げる必要は無いとする立場**である（結局，不都合を積極的に見つけ出すことはしないので，検討の範囲は，極値分布のみにとどまる）．他方，大から小，すなわち，**母分布の可能性となる風呂敷をいきなり大きく広げて，手当り次第に探す．その際に，極値分布がデータに「それなりに適合」していても，それで満足できずに，「より適合」する分布を「最適合な」分布と考えて，極値分布ではない分布を積極的に用いる立場である**．

また，母分布に含まれる**母数推定にも，一般的な統計理論（の常識）を重視するか否か**，すなわち，

名古屋の年最大降水量



Precipitation
log10 Return Period
Model 1901-2008
C.I. for the above
Model 1901-1999
0
100
200
300
400
−1
0
1
2
3
4



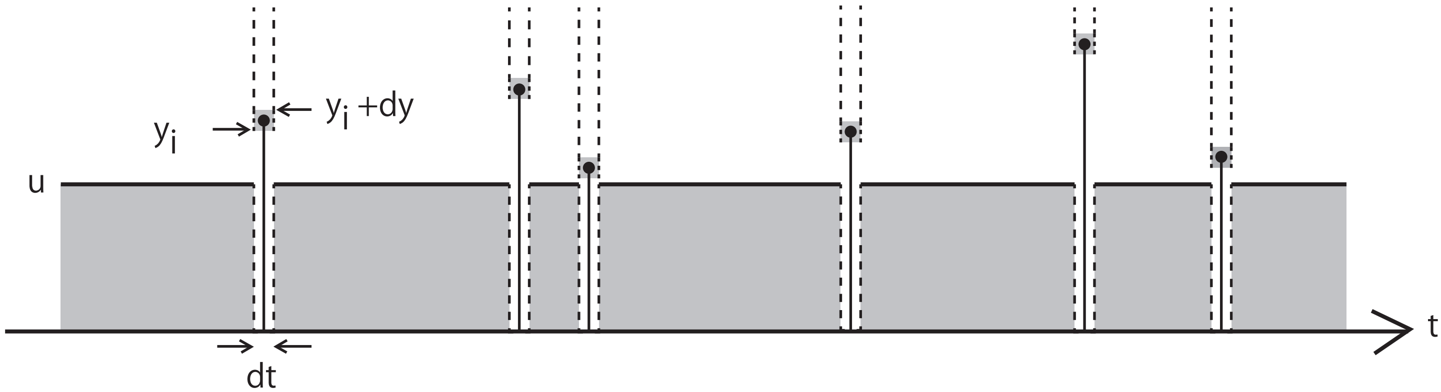
Precipitation
log10 Return Period
Model 1901-2008
C.I. for the above
Model 1901-1999
0
100
200
300
400
−1
0
1
2
3
4

また，母分布に含まれる母数推定にも，**一般的な統計理論（の常識）を重視するか否か**，すなわち，

考え方の相違２）パラメータ（母数）の推定には，**尤度を用いた最尤推定を用いるのが基本**である．最小自乗法を用いる回帰分析も，回帰係数の推定は，一般に最尤推定である（最小自乗法は最尤法を実行する際の計算のテクニックである）．最尤推定では都合が悪い場合に限り，別の推定法を模索する立場がある．これに対し，**最尤法に基づく統計理論には，推定誤差の漸近正規性が含まれており，その漸近近似の適用をあえて避けるならば，推定法にアレもコレも検討の余地があると考える立場**も生まれる．その場合に，小から大，すなわち，**最尤法の枠組みで考えて，それでは困ることが生じるから，徐々に検討の範囲を広げよう**と考えるか，あるいは，基本となる統計学における思考の枠組みを放棄して，あえて野放図に大きな範囲から考えを巡らせるかの相違がある．

あらゆることの初期段階では，試行錯誤を避けられず，そのために，過度に重厚な取扱いになるざるを得ない面がある．しかし，次第に全体像がつかめてくると，思考の節約も必要となる．すなわち，さらに向こうのことを考えるために，これまでの扱いについての枝葉末節を刈り込み，思考の見通しをよくするためには，**取扱いを簡素化することも重要になる．数学理論の多くに，極限操作を伴うのは，そのような簡素化のためである**と考える．**しかし，刈り込み過ぎも，現実離れしてしまう．**
例えば，流体運動の数理解析で，　. . .

点過程モデル
(Marked Point Process Model)



$$L(\boldsymbol{\theta}_1) = \left\{ \prod_{i=1}^{k} \frac{d\lambda}{dy}(y_i, \boldsymbol{\theta}_1) \right\} \exp\left\{-n\lambda(u, \boldsymbol{\theta}_1)\right\}$$

単なるカーブフィッティングではない

例えば，流体運動の数理解析で，渦無しの非圧縮性流体に対するポテンシャル理論の体系は美しいが，乱流の扱いができない点で，現実離れをしており，現実に生じる流体運動を把握する立場では決定的な問題があると言える．ただし，多くの問題では，乱流そのものを真正面から取組まなければばならない場合ばかりでなく，ポテンシャル理論をベースに，乱流が支配的となる領域では，境界層理論で補正することにより解決できる場合も多くある．このような「現実世界への対応」は，極値統計解析にも必要となると考える．また，極値統計解析を行うための数理的な理論体系も，現状ではまだ見通しが悪く，**多くの応用分野の人々の「往来」ができるような「道路整備」の必要性**を個人的には感じている．例えば，極値分布を特徴づけているのは，

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \approx G(z) \qquad \text{as} \quad n \to \infty$$

$$G^n(\alpha_n z + \beta_n) = G(z)$$

“non degenerate（非退化）” であり，これを “max stable（最大値安定性）” ともよんでいる．しかし，このような特徴が，どのような意味をもっていて，極値統計解析を行う上で，なぜこの性質が必要であるのか？そのような事情が，極値統計解析の利用者に十分に伝わっていないように感じている．

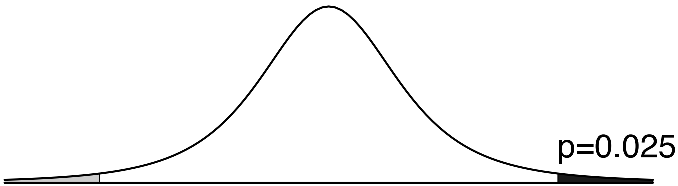
また，**データに分布関数を当てはめるフィッティングの問題として未だに認識**されていることも**たいへん残念**である．

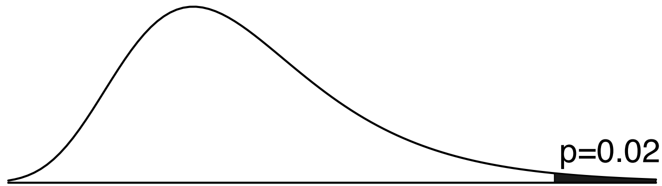
地上と天空をつなぐもの

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

ティ検定をはじめとして，正規標本に対する検定法が一般的に普及したのも，スネデカー・コクランによる「統計的方法」のような「よりよく使うための整理」があったからだと考える．なお，そこでは「自由度」という考えが極めて重要な役割を果たしているように思う．

　考えの相違が生じる理由の理解を深めるため，極値統計解析をとりまく背景を知る必要がある．
極値統計学は数理統計学の一部であるが，極値統計学が他と切り離された異質な統計学ではない．





そのため，極値統計を理解するためには，統計学のほぼ全体の知識を理解する覚悟も必要である．とはいうものの，対象となる現象についての深い理解も求められる技術者にとって，その余裕が無いのも現実である．そこで，どのような一般的な統計学の基礎知識が最低限に必要であるか？ 今回のワークショップでは，このことを整理することも，その目的の１つである．極値統計学の用語も含め，必要となる用語を挙げてみた．これらの用語を介せば，考え方の相違が生じる理由を深めることができると考える．また，現段階では，その相違を容易には解消できないかもしれないが，その先にある活用も考えてみたい．

極値統計学を理解するための用語集（案）

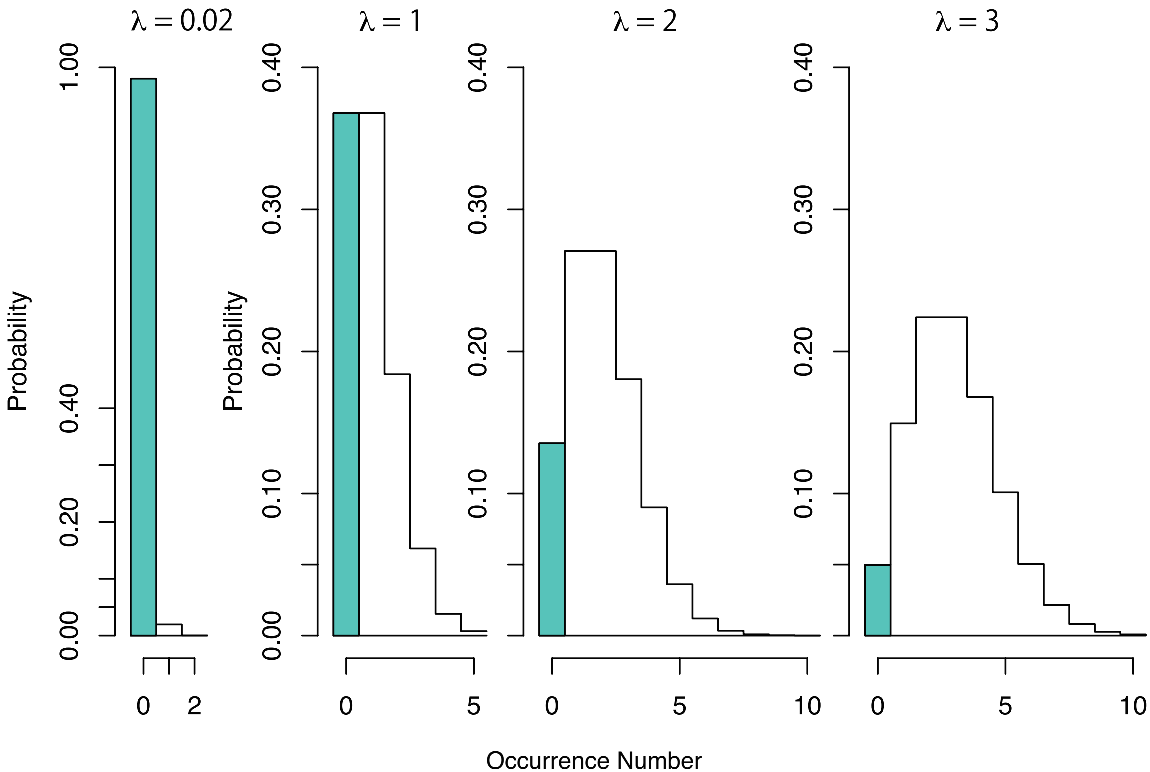
回帰と共変量（Covariate）
緩慢変動関数と正則変動関数
ガンマ分布（ポアソン分布の共役分布として）
区間推定と信頼係数
グンベル分布とガンベル分布，一般（化）極値分布
経験度と（正規標本に対する）自由度
再現レベル（確率外力）とクォンタイル
再現レベルの誤差分散（＝尺度母数の自乗／経験度）
母平均の推定誤差分散（＝母分散／標本サイズ）
再現期間（この逆数を超過確率，生起率？）
再生性（確率変数の和の分布の）
残差，誤差，標準誤差，標準偏差
指数分布族と十分統計量
順位統計量と順位相関係数
順序統計量と上位 r 番目までの極値分布
情報行列（Fisher 情報行列）
周波数と周期の関係（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 と Rule of Thirds

いずれもポアソン分布の性質



極値統計学だけに特化して勉強すると，単純に，理解困難に陥るだけでなく，誤解なども生じる可能性がある．

東京大学教養部統計学教室 編，基礎統計学，Ⅰ．統計学入門，Ⅱ．人文・社会科学の統計学，Ⅲ．自然科学の統計学，1991, 1994, 1992，東京大学出版会，308p., 404p., 366p.

また，本冊子では，統計学的および確率数学的な側面から高橋氏および志村氏による講義ノートを掲載しているが，さらに詳しく学ぼうと考える人のために，

1) An Introduction to Statistical Modeling of Extreme Values, by S. Coles, 2001, Springer, 209 p.

2) Extreme Value Distributions - Theory and Applications, by S. Kotz and S. Nadarajah, 2000, World Scientific, 196p.

3) Extreme Value Theory - An Introduction, by L. de Haan and A. Ferreira, 2006, Springer, 417p.

4) Order Statistics, 3rd Edition, by H. A. David and H. N. Nagaraja, 2003, Wiley, 480p.

5) Statistical Analysis of Extreme Values, 3rd Ed., R.-D. Reiss and M. Thomas, 2007, Birkhauser, 511p.

6) Statistics of Extremes: Theory and Applications, by J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers, 2004, Wiley, 514p.

7) 極値理論、信頼性、リスク管理，渋谷政昭・高橋倫也，21 世紀の統計科学，第２巻，第４章，pp.89-124.

8) 接合分布関数（コピュラ）の理論と応用，塚原英敦，21 世紀の統計科学，第３巻，第５章，pp.111-146.

`http://park.itc.u-tokyo.ac.jp/atstat/jss75shunen/`

また，ＩＳＭシリーズ：進化する統計数理（近代科学社）にて，極値統計学（高橋倫也・志村隆彰）の出版が今後に予定されている．