2016 年 6 月 17 日　統計数理研究所　オープンハウス

極値統計に伴うサプライズ！

統計数理研究所 客員 北野利一（名古屋工業大学）

９９年間，それほど大きな自然外力（たとえば，大雨）に見舞われることがなかったとしよう．そして，１００年め，未曾有の甚大外力の来襲を受けた時，『**過去に経験したことの無いような自然の脅威！まさか，来るとは思っていなかった**』という事態に直面する．その状況を視覚的に示すと，図 -1 のようになる（枚挙にいとまのないほどに数多くの例がある；最近では，2015 年インド南部豪雨など）．

質問： 過去に経験しない事象が生起したことでサプライズと感じるギリギリのタイミングとレベルに，シールを貼って示してください．

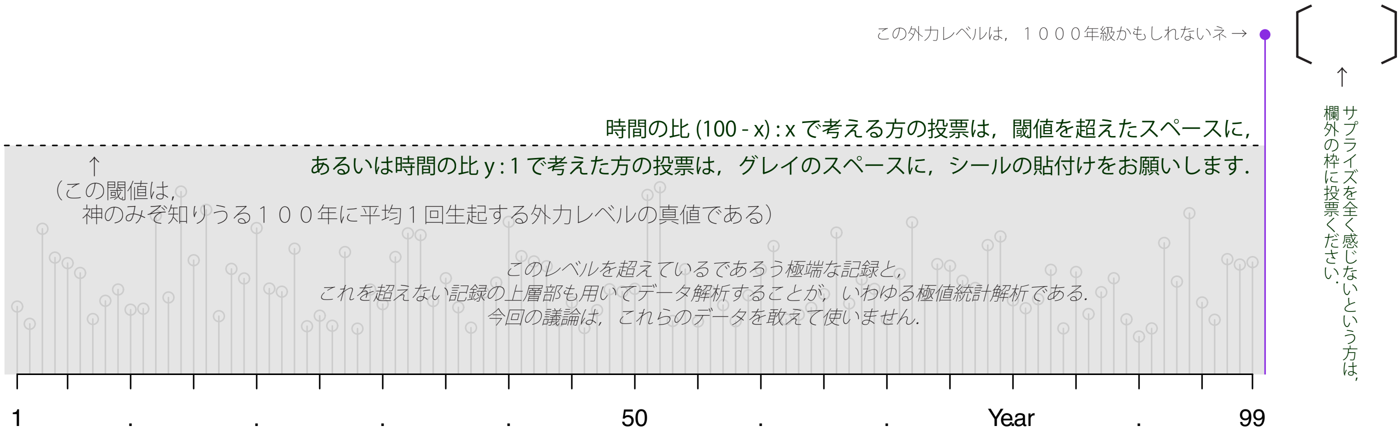


図 -1　それは偶然に生じたと解釈するからサプライズであるのか？／偶然に生じたとは解釈できないからサプライズなのか？

このような話は，以下のような**一種のパラドックスになっているのではないか？**というのが，ここでの狙いである（北野，2015a）．

1）事象の待ち時間分布に，$f(k) = (1-p)^{k-1}p$ で与えられる幾何分布を考える（p は事象が生じる確率）．この時，平均待ち時間（これを**再現期間**とよんでいる）は，$R = 1/p$ となる．ここで，$R = 100$ とすれば，100 年めで事象が生起する確率

$$100 = 1/p \qquad \rightarrow \qquad p = 0.01 \tag{1}$$

を得る．したがって，各年毎の事象の生起確率が希少（0.01）である場合に，99 年間に事象が生じずに最後の 100 年めに生じることは平均的な状況であり，この状況は，**偶然によるものだと考えている**ことになる．

2）99 年間を A 群，ひきつづく 1 年を B 群とおいて，〝**レベル超過の生起率を対象にした２標本問題**〟として考える．北野ら (2015) では，所与のレベルを超えない外力の情報を用いた極値統計解析の応用を示しているが，ここでは，その元となる**ポアソン検定**で考える．すなわち，A 群と B 群の生起数 $k_A = 0$ および $k_B = 1$ ならびに観測期間長の比 $N_A/N_B = 99$ に対して，両群の（単位時間あたりの）生起率に違いが無いとする**帰無仮説** $\lambda_A/\lambda_B = 1$ のもとでは，

$$q = \frac{N_A}{N_B}\frac{\lambda_A}{\lambda_B} \Big/ \left(1 + \frac{N_A}{N_B}\frac{\lambda_A}{\lambda_B}\right)\Bigg|_{N_A/N_B=99,\ \lambda_A/\lambda_B=1} = \frac{99}{1+99} \qquad \rightarrow \qquad 1-q = \frac{1}{100} \tag{2}$$

を用いて，２項分布により，p 値を求めることができる．**対立仮説**を $\lambda_A/\lambda_B < 1$ ととれば，

$$P(X_A \leq k_A \,|\, X = k_A + k_B) = P(X_B \geq k_B \,|\, X = k_A + k_B) = \sum_{X_A=0}^{0} \binom{1}{X_A} (1-q)^{1-X_A} q^{X_A} \Bigg|_{1-q=\frac{1}{100}} = 0.01 \tag{3}$$

となる．その値は比較的に小さく，帰無仮説を棄却して，既往の 99 年間と引き続く 1 年間において生起率の違いがあると判断する．すなわち，最後の 1 年は，それまでとは外力の生起の状況が異なり，**これまでの 99 年間に経験していない大きな外力が生じた（偶然ではなく必然的な理由がある）と判断**する．

以上の２つの異なる視点で，サプライズが解消できたと考えれば，**1）の偶然**と **2）の必然**で，**パラドックスになっている**とも言える．言い換えれば，**2）の必然をとるには，1）で偶然と扱う希少確率**を**アリエナイ確率**として扱っているから矛盾が生じるのだ．

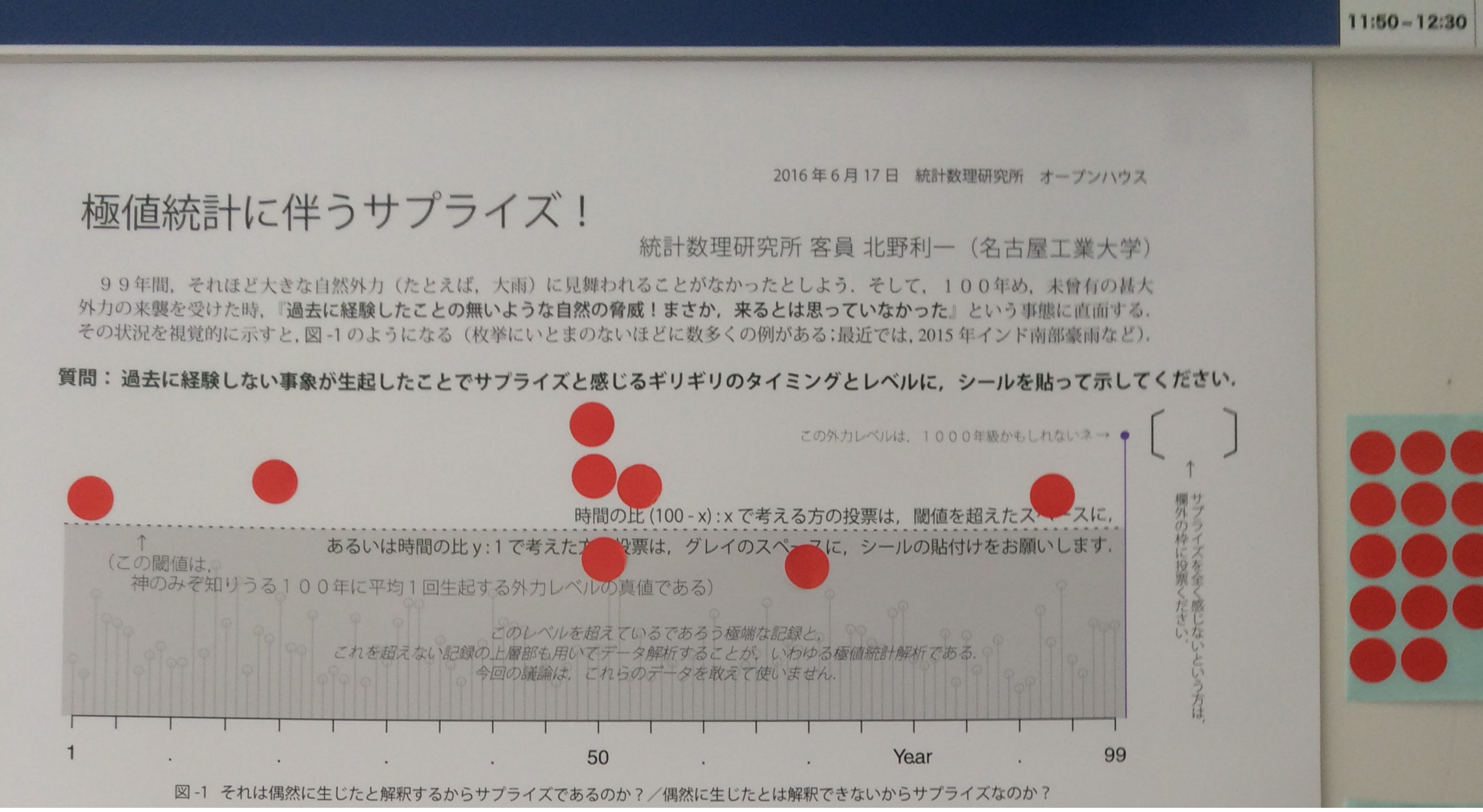
結局のところ，事が生じてから過去に振り返る際，**直前までの長い沈黙の時間との対比で，サプライズが生じる**のではなかろうか．すなわち，**１年というブロック**を用いるのは，ある意味において自然な取扱いであるかもしれないが，**この場合には適さない**のである．極端に言えば，１００年を５０年ずつの２ブロックに分割して，「前半の５０年には生じなかった甚大外力が，後半の５０年で生じた」と考えると，なにも驚く（サプライズ）に値しない．さらに，その外力の大きさが，５０年最大値のとりうる範囲から逸脱して，１０００年最大値がとりうるような値となっていても，前半１０００年のうち最後の５０年には観測されず，後半の１０００年のうち未だに５０年しか経過していない現時点で，そのような甚大外力がたまたま観測されたと理解することも，それほど現実離れ（サプライズ）したものではない．

さて，**ここで皆さんにお尋ねしたいのは，**沈黙の前半と事象が生じた後の後半の時間ブロック長の比がどの程度であれば，現実離れ（サプライズ）していると感じるのであろうか？ そのギリギリの状況として，**全体を１００年として前半と後半を分ける時点（事象の生起した時点）**に付属のシールで示していただきたいのである．例えば，８１年めに事象が生じて，前半８０年でゼロ回に対して，後半２０年に１回となることに違和感（サプライズ）を感じる方は，シールを８１年のところに１つ貼っていただきたい．なお，100 年を２分する (100 - x) : x を尋ねているのであるが，生起した後の区間は考えずに，y :1 を問題にしたい方もいるかもしれません．その場合はグレイの領域に印を貼って欲しい．いずれにせよ，ここで皆さんにご検討いただきたいのは，**極値の希少確率と，検定の有意水準のとり方の問題とも言える．**

今回とは逆に，観測開始して直ちに記録された外力が甚大過ぎて，それ以降１００年経過しても，新記録が出ないこともある．今回の話にも類似して，この場合もしっくりこないと感じる場合がある（1896 年の彦根豪雨）．これは次回に検討しよう．また，このようなパラドックスの他に，**「見逃し」と「空振り」の２つの過誤に対するジレンマ**も生じる（北野，2015b）．この検討も別の機会に譲る．

参考文献

北野利一 (2015a): R 年確率外力に対するパラドックス，統計数理研究所共同研究リポート，349, pp.37-41, 2015.
北野利一 (2015b): 温暖化適応策の意思決定 〜 ２つの誤りのはざまで，中部経済新聞，2015 年 2 月 20 日．
北野利一・高橋倫也・田中茂信 (2015): 降水量の極値特性の気候変動に伴う差異の検出 〜 変化の現れ方の想定により生じる問題，土木学会論文集 B1（水工学），第 71 巻，No. 4 pp. I_361-I_366.



投票していただいた方の結果は，上記のとおりである．左端の２つは，y:1 で考えた方の投票結果であるので，グレイの領域に貼っていただく方がよかったのだが，規模も考慮に入れたためか，グレイの外になったようである．

１００年間の最大値が特定の年に生じる確率は 1/100 であるが，その 1/100 の確率で生起するイベントが，どこに出現すると，サプライズを感じるのか？という人間の感覚を尋ねてみたわけである．もちろん，どこで生じても論理的には矛盾しない．また，０から１までの割合の数字で表される確率というのは，物理的な単位や次元の無い無名数であり，それゆえに，抽象化されている点で，数学的にいろんな展開が可能となることは，確率の有利な側面であるが，それが逆に，人間の感覚とのズレが生じる原因でもあると考える．そのため，希少確率の比較として，自然災害を引き起こす甚大外力の生起確率と，交通事故や殺人事件の遭遇確率などとの感覚の違いもそこに問題の所在があると考える．５０年あたりに印をつけられた複数の方からは，「やはり，人生の長さを代表スケールにとった感覚に基づくのかな？」とコメントをいただいた．

私にとっては，むしろ，欄外の「サプラズを感じない方」も少なからずいると思ったのだが，今回にポスターを訪問していただいた方には，該当者はいなかった．実は，私の問いの狙いには，「サプラズを感じない方」は，極値統計学の論理を素直に受け取っていただける方で，なんらかの時間スケールでサプライズを感じる方は，極値統計学の論理を頭で理解できても，感覚としては受入れ難いのかもしれないということにある．

すなわち，ここで言う**「感覚としては受入れ難い」極値統計学の論理**とは，例えば，年最大値データに対して，年最大値分布である一般化極値分布あるいはグンベル分布を適用して，母数推定を行う．そして，非常に小さな超過確率（例えば，1/100）となるクォンタイル点の値を推定することが正当化されていることである．

一般によく知られるティ検定は，母分布とする正規分布の平均を仮説で定めて，その仮説を棄却すべきか，容認すべきかを問う際に，有意水準（例えば，1/100）を定めて，その中に統計量が入ってくると，帰無仮説のもとでは，現実に生じにくい（アリエナイ！）ことになるので，むしろ，仮説が誤っていると考えて，帰無仮説を棄却するわけである．

ティ検定で用いる論理 と，**希少確率のクォンタイル点を推定する（推定できると肯定する）極値統計の論理** が相反するように感じるのである．私が問いたかったのは，このあたりの感覚の相違にあったのだが，今回の問いは，少し複雑（煩雑？）になったかもしれない．またの機会に，乞うご期待！（北野利一　記）