

電波研究の玉手箱

第3講 統計的にものを見る目を養おう

電気通信大学名誉教授
唐沢 好男

毎号、電波技術の基本テーマの中に見え隠れする不思議を玉手箱につめてお届けしている。第3講は、統計的なものの見方の話。信頼区間推定と極値統計を取り上げる。極値統計は想定外の出来事を想定外としないための心構えを教えてくれる。この視点に立って、我が国の降雨の年間最悪値統計を見てみたい。無線通信に影響が大きい降雨減衰、その雨の降り方に地球温暖化の影響が見えているのだろうか？



👉 想定外の出来事を想定外にしないために

大きな地震や津波、豪雨災害が起きたとき、想定外な事と驚く。新型コロナウイルスの蔓延もその先が見通せず不安な気持ちが拭えない。こういうときの判断に重要なのが、長期的な視点に立っての統計的なものの見方である(図1)。ここでは、観測データを客観的に分析するための、すなわち間違った結論を導かないための信頼区間推定と、想定外をもたらす最大値の性質を扱う極値統計を取り上げたい。そして、その統計的な目で我が国の降雨量年間最大値の長期統計について見てみたい。

降雨は、10GHz以上の高い周波数の電波に対して大きな減衰をもたらす無線伝送の大敵である。近年、

地球温暖化によって、様々なところの自然環境に異変が現われ、この問題への警鐘が鳴らされている。では、雨の量についてはどうであろうか？統計的にみてみよう(図2)*1。

👉 信頼区間指定：結論は慎重に

ある量 x と y に関して、 x が原因(説明変数)、 y が結果(観測値、目的変数)であったとする。このとき、本当の関係が $y = a + bx$ で与えられる線形関係の場合を考える。実際に観測される y には、 x とは別の要因によって発生する変動成分(誤差： e) が加わり、観測される目的変数の試行 i ごとの値は、次式となる。

$$y_i = a + bx_i + e_i \quad (1)$$

このとき、 x と y の組で与えられる n 個のデータ $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ の処理によって得られる回帰直線

$$\hat{y} = \hat{a} + \hat{b}x \quad (2)$$

は、母集団の特性 ($y = a + bx$) を推定するものになる。

回帰直線の決定にはガウスが編み出した最小二乗法が用いられる。最小二乗法では回帰直線と観測値との

*1 本内容の詳細や関連文献は技術レポート[1]にまとめている。本稿で、その十分な根拠を述べることはスペース的にできず、結果のみを示している。不明な点は[1]を見てほしい。

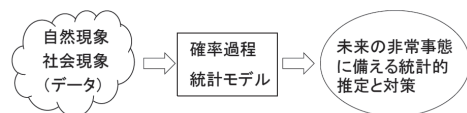


図1 想定外の出来事を想定外にしないために

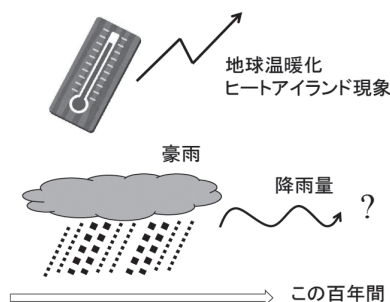


図2 気象変化の長期的傾向は？

ずれ $\varepsilon_i = y_i - \hat{y}_i$ の 2 乗和 (平均二乗残差) が最小になるよう $\hat{a}, \hat{b}$ を定める。回帰直線は、データが与えられれば一意に定められる (定め方は [2] の (49) 式に)。この回帰直線は、傾向を捕らえるのに分かりやすいが、どの程度信用できるのかという不安がある。

回帰直線の信頼性を定量的に評価するために生み出された概念が区間推定である。本当の特性が有りそうな範囲を上限と下限で囲み、信頼区間と呼ぶ。この確率の目安値を 90%, 95%, 99% などと定め、95% が良く用いられる。95% 信頼区間の場合、その範囲内にある全ての特性 (直線で表される) について 5% 以上の存在確率が有り、どれも棄却できない、とみなす。要は、「回帰直線だけで判断するのは危ないですよ、区間内に直線で引かれる傾向は、全部有りそうですよ」ということである。

具体的な信頼区間 (上限と下限) の定め方は割愛する ([2] の式 (61) に)。信頼区間の特徴を図 3 により説明する。同図では、真の特性 ($y=3+x$ を仮定) を点線①で、データ (●) に対する回帰直線を実細線②で、95% 信頼区間の上限と下限を実太線③で示している。図の回帰直線は x の増加に対して y も大きくなる傾向を示しているが、例えば一点鎖線④のように、 y が減少する特性も信頼区間の中に 5% 以上の確率で存在しており、このデータからは「 x が大きくなれば y も大きくなる」と言う答えを出せない、が区間推定の約束事 (ルール) になる。ただ、データ数を増やして行くと (この例では、例えばデータ数を 20 とすると)、信頼区間の幅が狭まり、「 x が大きくなれば y も大きくなる」という結論が言えるようになる。信頼区間推定は、傾向を積極的に見出そうとするものではなく、だめなもの (= 信頼区間外の特性) を棄却するという用心深さに比重を置く推定になっている。

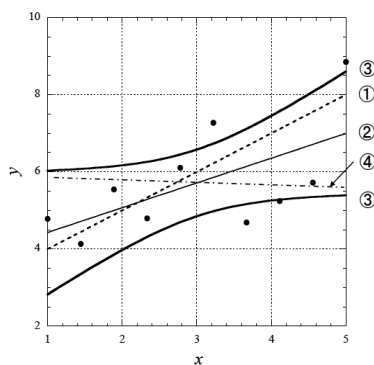


図 3 $y=3+x$ の特性①にばらつきを加えたデータ (●) の回帰直線②と 95% 信頼区間の上限と下限③、その中に含まれる一つの例④ (95% 信頼区間は、真の特性が、回帰直線を中心にしてこの範囲にある確率が 95% であるという意味であって、データのばらつき幅を表しているものではないことに注意)

極値統計：最大値には最大値特有の性質がある

グループ毎の極値 (通常は最大値) だけを集めて、その統計を扱うのが極値統計である。グループ内の分布がどのような確率分布であっても、極値だけ集めたものの分布は元の分布に依存しない極値データ特有の分布 (= 極値分布) になる性質が知られている [3]。

極値分布には三つのタイプがあるが、無線通信に良く出てくる正規分布やガンマ分布のように指数関数が積の形で入っている分布に対しては、グンベル (Gumbel) 分布と呼ばれる次式の形になる。

$$f_G(z) = \exp\{-z - \exp(-z)\} \quad (-\infty < z < \infty) \quad (3)$$

$$z = (x_{\max} - b_n) / a_n$$

ここで $x_{\max}$ は n 個のデータの最大値、 a_n, b_n は適当な規格化係数である。図 4 は、後で出てくる日本各地の日降水量年間最大値を、それぞれの地点の平年値で正規化したもののヒストグラムとそれに対応するグンベル分布を描いている。(3)式は不思議な形だが、実際の最大値データの分布とよく合っていることが分かる。

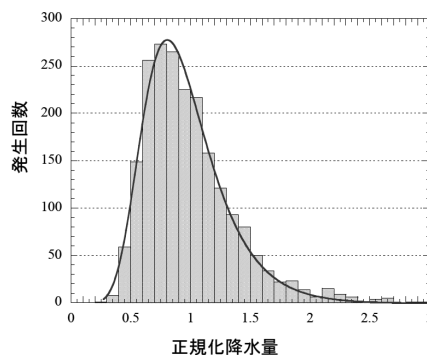


図 4 正規化年最大日降水量のヒストグラム (後述) と対応するグンベル分布 (実線)

地球温暖化問題

近年の気温の増加について、地球全体では $0.73^\circ\text{C} / 100$ 年の、日本 (都市化による影響が比較的小さい 15 地点) では $1.21^\circ\text{C} / 100$ 年の増加が報告されている [4]。大都市を代表する東京での年間平均気温の上昇率は $3.2^\circ\text{C} / 100$ 年のペースで、日本各地の中で一番大きい。東京 (大手町にある東京管区気象台エリア) は都市化率が日本の都市の中で最大 (92.9%) であり、地球温暖化に加え、都市構造の変化によるヒートアイランド現象も加わっているためと説明されている。図 5 は、100 年間 (1920-2019) での、東京の毎年の平

均気温を示している。区間推定によって、回帰直線の傾向が信頼できるものであることを示している。このように、ここ 100 年のスケールで見ると、確実に地球温暖化は進んでいる。では、降水量データで見るとどうだろうか？

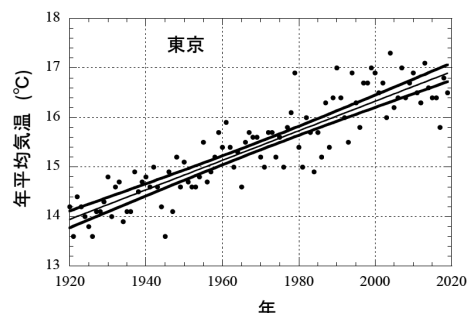


図5 東京の年平均気温の推移（細線：回帰直線、太線：95% 信頼区間）（地球温暖化+ヒートアイランド現象）

年間最大降水量の長期傾向を見てみよう

日本各地（約 1,300 箇所）の気象データが気象庁のホームページから公開されている [5]。降水量に関しては、一日、1 時間、10 分間に降る水の量、すなわち、日降水量、時間降水量、10 分間降水量について、観測開始から今日までのおよそ 100 年間のデータベースがある（期間は観測点やデータの種類によって異なる）。年間の最大値を与える降水現象は、例外なく降雨であるので、本文中での降水量は降雨量と読み替えてよい。

気温上昇率が我が国で一番高い東京について、時間降水量の年間最大値を図 6 に示す。回帰直線を見ると僅かな上昇が認められるが、信頼区間推定では、温暖化の影響が現われているとは言えないことになる。文献 [1] で、筆者が調べた限りにおいて、実は、東京・大阪の大都市ばかりでなく、我が国の広い範囲に亘って、降水量（日・時間）の年間最大値には、地球温暖化の影響はまだ見えてこなかった。

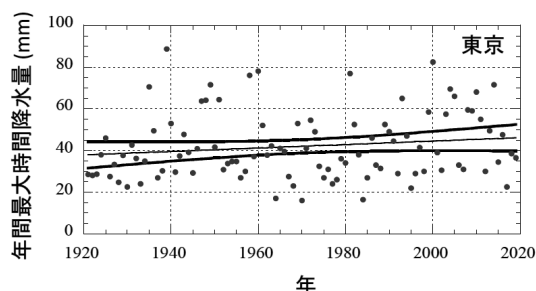


図6 東京における時間降水量の年間最大値の年変化（年毎のばらつきが大きくはっきりした傾向は見えていない）

大災害を引き起こす尺度になるであろう日降水量についても、土砂降りや降雨減衰量に影響が大きい時間降水量についても、信頼区間推定では、地球温暖化の影響が認められず、筆者には意外な結果であった。

極値統計から見る千年に一回の豪雨とは？

前節の評価において、各種降水量の年間最大値に有意な長期的傾向が見られなかったため、これが定常状態としてさらに長く続くとした場合、その状態での 100 年、200 年、1000 年に一度の大雨、すなわち、そのような期間で発生する最大値はどのくらいかを調べてみた。現実には、100 年程度のデータしかないため、気象現象において独立とみなされる程度に離れた多数の点（目安としては 100km 程度以上の距離にある仙台以南の 18 地点 [1]）のデータを集めて数を増やし、以下の手順で最大値の特性を調べた。

- ①各地点の年間最大値データを集める
- ②地点毎に年間最大降水量の平均値（＝平年値）で正規化した降水量（＝平年値を 1 とする正規化最大降水量）を求める
- ③②のデータを全部まとめて、仮想 1 箇所のデータにする。（データ数は 2000 前後）
- ④ N 年間の最大値の平均値を求める。以後この N を再現年数と呼ぶ。具体的には、全データを N 個単位で分け、それらの最大値の平均を求める。
- ⑤③の段階での日降水量の年間最大値（正規化されたもの）全データのヒストグラムを前述の図 4 に示している。図より、極値分布であるゲンベル分布によく合っており、極値統計の適用が期待できる。

図 7 は、三つの正規化降水量（日・時間・10 分間）を再現年数 N に対して示している。横軸の N を 3000 年のスケールまで表していて、仮想年数である。図より、3 つの特性は、どれも直線形状になっており *2、再現年数 N と最大降水量期待値 $\langle R_{\max}(N) \rangle$ の関係は次式で近似できる。

$$\langle R_{\max}(N) \rangle \approx (\alpha \log_{10} N + 1) \langle R_{\max}(1) \rangle \quad (4)$$

$$\alpha = \begin{cases} 0.867 & \text{(日降水量)} \\ 0.659 & \text{(時間降水量)} \\ 0.510 & \text{(10分間降水量)} \end{cases}$$

（注：[1] では日降水量を $\alpha = 0.909$ としているが、0.867 に修正）

*2 降雨強度の確率分布はガンマ分布で近似できることが調べられている。極値統計理論によると、ガンマ分布する量の極値と再現期間との関係は対数座標上ではほぼ直線的になることが導かれている [3]。故に、図 7 が示す直線形状は、極値統計理論に符合している。

10 年、100 年、1000 年に一回の大雨は、年間の最大値平均 $\langle R_{max}(1) \rangle$ に比べて、日降水量では 1.9, 2.7, 3.6 倍、時間降水量では、1.7, 2.3, 3.0 倍、10 分間降水量では、1.5, 2.0, 2.5 倍である。東京の日降水量年最大の平均値は 119.2mm、10 年・100 年・1000 年に対しては、226mm・322mm・429mm に、時間降水量年最大の平均値は 42.06mm、10 年・100 年・1000 年に対しては、72mm・97mm・126mm になる。驚くべき数値かもしれないが、極値統計はそれを予想している。

一つの地点で見れば、1000 年に一回の大惨事も、気象的に独立な 50 地点（パラレルワールド）を俯瞰してみれば、エリア全体としては、20 年に一回、どこかで起きる最悪値と読み替えることもできる（図 8）。

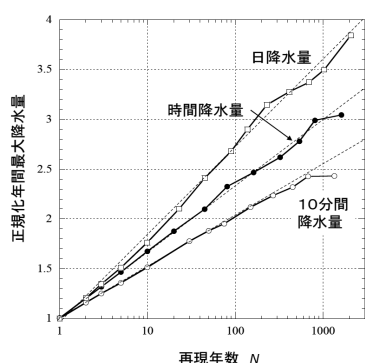


図7 N 年間での最悪値推定（点線は $N=1\sim500$ の部分の対数回帰直線を $N=3000$ まで外挿）

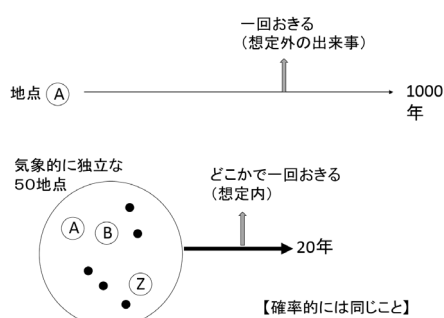


図8 想定外の出来事もパラレルワールドで見ると

👉 ものの見方

統計的にものを見ることの大切さとして、信頼区間推定と極値統計を紹介した。前者は、少数のばらつきあるデータから間違った答えを導かないことに、後者は、想定外と思われる出来事も想定内にすることに、知見を与えてくれる。

上記二つの統計の応用として、降水量の年間最悪値に着目したのは、①近年、記録的な水害（川の氾濫や

土砂崩れ）が毎年のように起こり、地球温暖化問題との関連で語られていること、②筆者自身、職業柄、降雨現象（降雨強度や降雨減衰）に強い関心があったこと、そして、一番肝腎なことであるが、③解析を可能とする気象データが気象庁 URL から公開されていること、による。

解析を通じて、年間最大値に着目した降水量に限ってみれば、世間で言われているほどには、地球温暖化の影響、すなわち、異常気象の長期傾向は見えてこなかった。何か見えるはずと思って着手したが、勘が外れたというのが正直な気持ち。注意してほしいのは、この結論は、ここで扱ったデータや処理方法の範囲でと言う条件付であり、日本の降雨には、地球温暖化の影響は現れていない、と言う一般的な結論を述べているわけではない。

一方、日本全体で見た降水量の年間最大値データは全てを一つの確率現象で捕らえる極値統計学の土俵には、極めて良く乗っていることが分かる。この視点に立って、このまま、同じ確率分布に支配される世界が長く続いたらどうなるかという予想をしてみた。東京において、100 年に一回の日降水量の最大値を 322mm と推定した（実際、この 144 年間（1876-2019）では、300mm 以上は 1958 年の 371.9mm の一回だけ）。この値をどう思うかである。生きている間に経験するかどうかと言う値であるから、東京に住み続けている限り実際に会うことはないかもしれない。しかし、パラレルワールドと言う考え方がある [6]。一箇所に留まってそこだけを見ていれば、そういうことだろうけれども、例えば、広いエリア全体で 20 の独立な気象が現れていれば、そのエリアのどこかには 5 年に一回、100 年に一回の大雨が降ることになる。テレビのニュースで知ることはこの出来事である。

本稿では、特に後半部分の降水量最悪値統計について、少ないスペースに多くのことを圧縮して収めている。冒頭脚注で述べたように同内容の詳細を技術レポート [1] にまとめているので、地球温暖化問題と降雨の関係に興味をもたれた皆さんは是非そちらもみていただきたい。

<参考文献>

- [1]-[3] の技術レポートは以下の唐沢研究室 URL から公開している。
<http://www.radio3.ee.uec.ac.jp/report.htm>
- [1] 唐沢好男, YK-021-rev, [2] YK-019, [3] YK-020
- [4] 気象庁, “地球温暖化予測情報 第9巻,” 2017.03, <https://www.data.jma.go.jp/cpdinfo/GWP/Vol9/pdf/all.pdf>
- [5] 気象庁, 過去の気象データ検索, 気象庁ホームページ, <https://www.data.jma.go.jp/obd/stats/etrn/index.php>
- [6] 藤部文昭, 統計から見た気象の世界、気象ブックス 041、盛山堂書店、2014。