

確率限界法検定を基礎とした極値降雨量の統計的解析

A Statistical Analysis on Extreme Rainfall Based on Probability Limit Method Test

都市人間環境学専攻 清水啓太
Keita SHIMIZU

1. 研究の背景及び目的

従来の洪水対策では、過去の降雨観測により蓄積された極値降雨量の標本がある確率分布に従い生じたものと仮定し、その確率分布において、計画超過確率（計画確率年）に相当する降雨量値を計画降雨量として設定してきた。計画降雨量とは、河川計画において設計基準となる降雨量値であり、計画降雨量が多らず洪水外力を防御・低減させるために、堤防高、遊水池および洪水防御施設などの検討が行われる。ここで、極値降雨量の標本としては、降雨量の観測時系列のうち年最大値から構成されるものが多く用いられ、この標本は年最大値資料という。また、上述した統計解析手法は、水文頻度解析と呼ばれる。水文頻度解析手法の多くは、極値理論にその数理的基礎を置く。極値理論とは、標本中の最大値は、サンプル数が多いほど、漸近的に二重指数型の関数形をもつ極値分布に収束することを示したものである。一般的に、計画降雨量を算定する水文頻度解析は、次に示すような手順を経て行われる。まず、頻度解析に使用する確率分布をいくつか候補として挙げ、次に、これらの確率分布を極値降雨資料に当てはめ、母数推定を行う。それらの推定確率分布群に対して、観測データとの適合度および、推定確率降雨量の安定性を評価し、最良の分布形の一つを採用する。当該の唯一つの確率分布において、計画超過確率に相当する降雨量値を計画降雨量として定める。我が国の一級河川流域では、100年から200年といった計画確率年に相当する降雨量値が河川整備における基準値として採用される場合が多い。ここで、年最大値資料を用いる際、 T 年確率降雨量（対象超過確率を p とした際、採用した確率分布の $100(1-p)\%$ ile値）は、1年間で当該値を超過する降雨の生起確率が p 存在することを意味する。このとき、確率年 T は対象超過確率 p の逆数として定義される。しかしながら、我が国における降雨の観測期間は、一級河川流域においても数十年間、多くとも百数十年間分程度であり、計画確率年で定められた期間よりも短いため、今日までに蓄積されている極値降雨量の観測値（観測極値）の総数に基づく、計画降雨量の推定は十分な精度とは言えない。すなわち、年最大値資料を使用する場合、その観測値は数十個から多くとも百数十個程度であるため、従来の水文頻度解析における推定は不確実性（推定誤差）を大いに含むといえる。加えて、近年における記録的豪雨の多くは、計画降雨量を超過するものであり、今後地球温暖化に伴う気候変動により豪雨の規模は増加するという予測が示されている。観測極値の不足は、このような記録的豪雨の統計的評価を困難とする。具体的には、観測極値の不足に起因する問題点として、記録的豪雨の観測値の多くが、採用した確率分布モデルから大きく外れ、数千年あるいは数万年といった、計画確率年に基づく河川整備では対応不可能な再現期間（確率年）を示すことや、データの観測年数を超える長期の再現期間に相当する確率降雨量を推定する際、確率分布モデルによる外挿に伴い、推定誤差が増大していくことがある。また、従来の水文頻度解析は、降雨量の定常性を仮定している。ここで、定常性とは、降雨量を支配する確率法則が時間的に変化しないことである。従来手法では、降雨量が従う確率分布を時間的に変化しないものとして扱ってきた。しかしながら、近年では、地球温暖化に伴う気候変動の影響による降雨量の非定常性が指摘されてきている。非定常性とは、降雨量の時系列がトレンドや周期をもつことあるいはそれらが従う確率分布が時間的に変化することをいう。観測情報の不足ならびに、気候変動に伴う降雨の激甚化を考慮すると、計画降雨量の不確実性の定量評価や将来生起しうる大規模豪雨の予測手法の構築は、今後の洪水対策の検討において喫緊の課題となる。

本研究では、計画降雨量に代表されるような推定値がもつ不確実性を定量化する信頼区間ならびに、将来の観測値がとり得る範囲を表す予測区間を構成する手法を構築した。信頼区間とは、有限な観測データを基に、母集団分布の未知母数 θ （ T 年確率降雨量値など）を推定するとき、下側信頼限界値 $L_{C.I.}(X)$ 、上側信頼限界値 $U_{C.I.}(X)$ を求め、これらの信頼限界値から構成される区間 $[L_{C.I.}(X), U_{C.I.}(X)]$ の中に、母数 θ を $(1-p)$ の確率で含むものである。ここで、 p は有意水準である（ $0 < p < 1$ ）。このとき、区間 $[L_{C.I.}(X), U_{C.I.}(X)]$ を $100(1-p)\%$ 信頼区間

という。予測区間とは、将来の観測値 Y が存在する範囲である。100(1- p)%予測区間は、下側予測限界値 $L_{p,l}(X)$ と上側予測限界値 $U_{p,l}(X)$ から構成される区間 $[L_{p,l}(X), U_{p,l}(X)]$ の中に(1- p)の確率で将来の観測値 Y の実現値が含まれることを意味している。また、本研究では、当該区間を構成する際、極値降雨量の母集団分布として、仮定した分布の下で、発生しうる豪雨の標本を棄却限界値として、導出可能とする確率限界法検定の理論を採用した。加えて、降雨量の非定常性を頻度解析に反映させ、将来気候下の確率降雨量を推定するためには、これまでの観測実績に加え、気候変動予測情報を取り込む必要があることを鑑み、大量のアンサンブル気候シミュレーション結果を有するd4PDFにおける計算降雨量値群をベイズ手法により取り込むことで、将来気候下での年最大降雨量に関する極値分布の推定手法を構築した。さらに、その発展形として、過去のデータを基に、長期先の予測を行う際、確率降雨量の予測精度が劣化していくという実態を統計的情報量の観点からモデル化し、その推定不確実性の増大を信頼区間・予測区間により定量化可能な手法を構築した。

2. 本論文の構成

本論文は、全8章で構成される。各章の内容と成果の概要は、以下のとおりである。

第1章の「序論」では、水文頻度解析およびその数理的基礎である極値統計理論の歴史を概説するとともに、利用可能な観測極値の不足に起因する、水文頻度解析上の問題点を指摘した。

第2章の「確率限界法検定の理論」では、同検定理論を詳説した。同章では、理論の背景にあるKolmogorov-Smirnov検定の概説をするとともに、両検定の検定力を比較し、確率限界法検定の水文頻度解析に対する有用性を示した。

第3章の「確率限界法検定に基づく信頼区間・予測区間の構成手法と当該区間を導入した水文頻度解析」では、確率限界法検定に基づいた信頼区間・予測区間の構成手法を詳説するとともに、当該区間を導入した水文頻度解析手法を提示した。

第4章の「気候変動予測情報を用いた極値降雨量に関する信頼区間・予測区間の将来変化の推定」では、ベイズ統計理論であるマルコフ連鎖モンテカルロ法により、気候変動予測結果を取り込み、地球温暖化時における信頼区間・予測区間の推定手法を構築した。

第5章の「全球平均気温上昇シナリオ別の極値水文量の将来予測」では、マルコフ連鎖モンテカルロ法を用いた、将来気候における年最大降雨量の極値分布ならびに、その信頼区間・予測区間の推定手法を我が国の経済活動上重要となる利根川流域に適用し、地球温暖化シナリオごとに、当該区間の将来変化を算定した。さらに、将来気候下における最大クラスの降雨パターンやそれらがもたらす洪水ピーク流量値を明示した。

第6章の「経験度・観測情報の耐久性を考慮した極値降雨量の統計的予測」では、経験度と耐久性の理論に基づき、将来期間における確率降雨量の推定不確実性が増大する過程を定量化し、それから算出される耐久性の値を用いて、確率限界法検定に基づく信頼区間の時間発展の算定手法を示した。

第7章の「観測情報と気候変動予測を統合した極値降雨量に関する信頼区間・予測区間の将来変化の推定」では、確率降雨量の予測精度の経年劣化と極値分布の母数に関するFisher情報量の減衰過程の関係を数値化し、時間軸を踏まえた、計画規模相当降雨量の不確実性評価手法を提示するとともに、その洪水対策への適用可能性を示した。

第8章の「結論」では、総括として、本論文の結果をまとめた。

3. 確率限界法検定の理論（第2章）

確率限界法検定は、Kolmogorov-Smirnov検定の弱点である“仮定した確率分布の裾部に対する検定力の低さ”を改善した検定理論である。ここで、検定力とは、誤った仮説を棄却する確率を表し、検定力が高いほど採択域（設定した有意水準の下で許容可能な観測値の範囲）は狭まり、推定の精度が向上する。同検定では、仮定した確率分布の確率表現関数の両側に確率限界線（確率限界法検定における棄却限界線）を構成し、観測値が両線により定まる区間の外部に存在する場合、有意な差があると判定する。ここで、確率表現関数 $\chi_X(u)$ とは累積分布関数 $F_X(x)$ の逆関数である。累積確率 $U(=F_X(X))$ は、区間 $[0,1]$ の一様分布（以降、標準一様分布あるいは $U[0,1]$ と表記）に従う。また、 U の実現値から構成された順序統計量 $\{u_{(1)}, u_{(2)}, \dots, u_{(n)}\}$ に関して、第 i 順序統計量 $U_{(i)}$ の確率分布は母数 $(i, n-i+1)$ のベータ分布となる。このため、式(1)が成立する。

$$F_{U(i)}(i) = P(U_{(i)} \leq u) \quad (1)$$

$$= I_u(i, n-i+1)$$

ここに、 $F_{U(i)}(u)$: 第*i*順序統計量 $U_{(i)}$ の累積分布関数、 $I_u(i, n-i+1)$: 母数 $(i, n-i+1)$ のベータ分布の累積分布関数、 n : 観測値の総数である。ここで、確率 α を「確率限界値を求めるための確率（極端な $U_{(i)}$ が生起する確率、すなわち、 $U_{(i)}$ が母数 $(i, n-i+1)$ のベータ分布の裾部付近に生起する確率）」、 $F_{U(i)}(u)=\alpha$ の解 u を「標準一様分布における下方確率限界値 $z_L(i)$ 」および $F_{U(i)}(u)=1-\alpha$ の解 u を「標準一様分布における上方確率限界値 $z_U(i)$ 」として定義する⁷⁾。すなわち、 $z_L(i)$ は母数 $(i, n-i+1)$ のベータ分布の100 α %ile値、 $z_U(i)$ は母数 $(i, n-i+1)$ のベータ分布の100(1- α)% ile値である。また、確率 α_{min} は式(2)により定義される。

$$\alpha_{min} = \min_{1 \leq i \leq n} \left\{ \text{Min} \left[I_u(i, n-i+1), I_{1-u}(i, n-i+1) \right]_{u=U_{(i)}} \right\} \quad (2)$$

ここに、 $I_u(i, n-i+1)|_{u=U_{(i)}}$: 第*i*順序統計量 $u_{(i)}$ の非超過確率、 $I_{1-u}(n-i+1, i)|_{u=U_{(i)}}$: 第*i*順序統計量 $u_{(i)}$ の超過確率である。式(2)は、「順序統計量 $\{u_{(1)}, u_{(2)}, \dots, u_{(n)}\}$ が有する非超過確率および超過確率をそれぞれ比較し、どちらか小さい方を抽出することで、 n 個の確率を得る。このとき、求めた n 個の確率のうち、最小値を α_{min} とする。」ということを表す。ここで、前述した n 個の確率を、 $\{\alpha'_1, \alpha'_2, \dots, \alpha'_n\}$ と記述すると、 $\alpha_{min} = \text{Min}\{\alpha'_1, \alpha'_2, \dots, \alpha'_n\}$ である。観測降雨量の標本は、実測された唯一のものしか得られない一方、その*i*番目の観測順序統計量の累積確率 $U_{(i)}$ は、ベータ分布に従い、これらのベータ分布は、過去観測する可能性のあった、観測第*i*順序統計量の生起確率の分布である。つまり、観測される可能性のあった観測標本自体を直接得ることはできないが、その累積確率は、標準一様分布から無制限にサンプリングすることができ、そのうち第*i*順序統計量 $u_{(i)}$ がベータ分布の裾野に発生した場合、その $u_{(i)}$ は、採用分布から大きく外れるデータ（極端豪雨の観測値）の生起確率として解釈可能である。このような $U_{(i)}$ の発生確率が α_{min} として表される。次に、確率限界法検定では、 α_{min} の分布を構成し、有意水準に応じて採用した α_{min} を確率 α として、確率 α および $(1-\alpha)$ に相当するベータ分布のクオンタイル値である、下方確率限界値 $z_L(i)$ および上方確率限界値 $z_U(i)$ を求める。次に、採用確率分布 $D(X; \theta)$ における確率限界値の導出手順を示す。 U は標準一様分布に従うため、標準一様分布における確率限界値を累積確率として扱うことができる。したがって、確率分布 $D(X; \theta)$ における下方確率限界値 $\chi_X(z_L(i))$ および上方確率限界値 $\chi_X(z_U(i))$ は、標準一様分布における $z_L(i)$ および $z_U(i)$ を採用分布の表現関数 $\chi_X(u)$ により、変換することで得られる。以上が確率限界法検定の理論である。確率限界値群に対して、採用分布と同一の関数を当てはめることで信頼区間、最良の適合度を示す分布形を算定することで予測区間が構成される。このように、当該区間の構成に必要な仮定は、降雨データの定常性ならびに降雨量の上限・下限を設定しないという、従来の水文頻度解析に一般的に用いられるもののみである。他の手法では、確率降雨量などの推定値に確率分布を仮定するパラメトリックな手法に基づく信頼区間・予測区間が多く提案されている一方、本研究はほとんど解析的に導出可能な信頼区間・予測区間を提示しており、既存の手法に対しても有用性をもつと考える。

4. 確率限界法検定に基づく信頼区間・予測区間に基づく水文頻度解析手法（第3章）

信頼区間の導入により、確率降雨量を取りうる値が理論的に推定可能となり、従来採用してきた観測実績に基づく*T*年確率降雨量値はその分布の最頻値となる。従来では、上述した最頻値を計画降雨量として採用してきたが、このような従来手法は、記録的豪雨の確率年を合理的に評価できないという問題をもつ。信頼区間を考慮することで、*T*年確率規模の洪水対策において、*T*年確率降雨量の振れ幅が求まり、上側信頼限界値に代表される記録的豪雨をその振れ幅の中の現象として解釈され、記録的豪雨の確率年が合理的に評価される。具体的には、上側信頼限界値の超過確率は、「対象確率年*T*」と「信頼区間の片側確率 $p/2$ 」との積により求まるため、*T*年確率の洪水対策において、100(1- p)%上側信頼限界値を超える降雨量が発生するリスクは、 $(1/T) \times p/2$ により表すことができる。このように、信頼区間の導入により、従来では考慮されなかった規模の激甚豪雨が生起するリスクを定量的に求めることが可能となる。加えて、上記の枠組みにより算出されるリスクは他分野におけるリスク（例えば、交通事故で死亡する割合等）と同程度のオーダーとなる。すなわち、激甚豪雨が生起するリスクを他分野におけるリスクと比較・評価することができるようになる。次に、名古屋地方気象台における99年間分の年最大日降水量の観測値、これら99個の観測値に当てはめたGumbel分布、東海豪雨（2000年）の観測値を対象として、確率限界法検定に基づく99%予測区間を用いた水文頻度解析を実

施した。予測区間の導入により、従来では“想定外”として扱われるような規模の豪雨を100年あるいは200年といった計画規模相当の確率年で評価できる。従来の枠組みでは、その確率年が計画規模を大きく逸脱するものとして扱われる東海豪雨の観測値は、99%上側予測限界値と新たに解釈され、かつ、当該豪雨の確率年は200年となる。加えて、激甚豪雨が将来期間において発生するリスクを予測区間により定量化できる。すなわち、上述したリスク（上側予測限界値の超過確率）は、「対象確率年 T 」と「予測区間の片側確率 $p/2$ 」との積により表される。東海豪雨の観測値は、200年確率年最大日降水量の99%上側予測限界値となるため、当該値の超過確率は、上記したリスクの算出式より、 $(1/200) \times 0.005 = 1/40000$ となる。すなわち、東海豪雨級の激甚豪雨が将来生起するリスクは1/4万となる。このように、この統計的予測手法の枠組みは、甚大な被害を将来もたらしうる豪雨の規模とその生起リスクを把握可能とするものである。

5. 地球温暖化進行時における極値水量の将来予測（第4章・第5章）

d4PDFへの力学的ダウンスケーリング結果に基づく確率降雨量の頻度分布と、確率限界法検定に基づく信頼区間の比較を行い、両者が極めて高い精度で一致することを示した。さらに、マルコフ連鎖モンテカルロ法により、当該のダウンスケーリングデータにおける将来予測値を取り込み算出された、信頼区間についても、確率限界法検定による理論と整合した。この結果は、力学的ダウンスケーリング計算の妥当性を理論的に支持するものである。この結果により、今後の気候変動を踏まえた洪水対策の検討にあたり、気候変動予測情報の導入可能性が示された。次に、利根川上流域を対象に、降雨流出モデルを用いて、温暖化進行時における降雨の将来変化が極値河川流量に及ぼす影響を推定するとともに、これらの極値河川流量をもたらす大規模豪雨の分析を実施した。その結果、産業革命の頃から全球平均気温が4°C上昇した将来気候下（RCP8.5シナリオ）における最大のピーク流量値は29650.6[m³/s]と算定された。この値は、現行の基本高水である22,000[m³/s]を大幅に超過するものである。加えて、このような大規模な洪水流量値をもたらす気候場は、太平洋高気圧が日本側に張り出した気圧場において、台風が太平洋沖から関東地方に接近・上陸し、太平洋沖に移動していく、という形態であることを示した。

6. 観測情報と気候変動予測を統合した極値降雨量に関する信頼区間・予測区間の将来変化の推定（第6章・第7章）

地球温暖化に対応した洪水対策を気候変動のモニタリング等を活用し、段階的に整備していく際に、何年先には計画降雨量の不確実性はどの程度になるか、ということを実量的に示すことは重要となる。これを実現するためには、これまでに蓄積されてきた観測値がもつ情報量を用いて、計画確率規模相当の確率降雨量値を推定する際、時間経過に伴い、当該の確率降雨量値に対する推定精度が劣化していく過程を数理的にモデル化する必要がある。本研究の最終的な成果として、観測と気候変動予測情報の双方を考慮した、確率降雨量ならびにその信頼区間・予測区間の予測手法を示した。本手法の有意性は、過去のデータを基に、長期先の予測を行う際、確率降雨量の予測精度が劣化していくという実態を統計的情報量の観点からモデル化し、その推定不確実性の増大を信頼区間・予測区間により定量化した点にある。また、d4PDFによる気候変動予測情報は、温暖化シナリオ別に年最大降雨量の時系列を有しているため、これらの時系列をマルコフ連鎖モンテカルロ法により、逐次的に取り込むことで、任意の将来時点における確率降雨量の予測値が算定可能となる。なお、本手法では、観測期間中に得られた各データがもつ情報量を全て等しいと仮定し、その仮定の下では、観測値時系列が定常過程であるという想定をすることが自然と考えるため、母数の定常成分が有するFisher情報行列の大きさ（ノルム値）を観測値の総数で除した値を観測値1個分がもつ情報量として扱っている。そのうえで、将来期間では、時間経過に伴う、母数の推定不確実性の増大を、予測に利用可能なデータ数として表現された観測情報の減少と結びつけている。確率限界法に基づく信頼区間・予測区間の幅は、使用するデータの総数に応じて決定するため、将来期間において残存する情報量を定量化することで、当該区間の将来変化予測が可能となる。この結果により、想定される温暖化シナリオのもとで、段階的な洪水対策の整備を検討する場合、本提案手法により数値化される、計画降雨量の逐次的な将来変化は、現実の時間軸を踏まえた事業オプションを検討する際の有益な指標となると考える。

7. 結論（第8章）

第8章では、各章の成果をまとめ、総括とした。

