

月最大降雨量の確率密度関数

芝浦工業大学 システム理工学部 数理科学科 2 年
BV24029 白金璃空

2025 年 11 月 4 日

目次

1	はじめに	2
2	降雨データとデータの性質	2
2.1	降雨資料の調査年数・調査年数	2
2.2	降雨ごとの確率分布	2
3	確率密度関数	4
3.1	採用した確率密度関数	4
3.2	適合性の検討	5
4	確率密度関数の統合化	8
4.1	降雨の基準化	8
4.2	数正規分布による確率雨量の統合化	8
4.3	最小二乗法	9
4.4	統一化の結果	9
5	終わりに	10

1 はじめに

今回の研究では、日本で長期間の降雨資料を得られている 6 地点を選び、月ごとの最大降雨量の確率密度関数を検討した。そして、より適合度の良い関数を抽出すると共に、降雨量の基準化によって、降雨の時間・場所に関わらず適用できる関数を導出した。

2 降雨データとデータの性質

2.1 降雨資料の調査年数・調査年数

研究で採用した降雨資料は気象庁の各観測地点のデータから入手したもので、博多・神戸・名古屋・さいたま・仙台・札幌の 6 地点で行った。また、月最大降雨量は 2003 年 1 月から 2024 年 3 月の 21 年間の資料を用いた。

2.2 降雨ごとの確率分布

次に、月最大降雨量の確率分布を、各地点からまとめた。

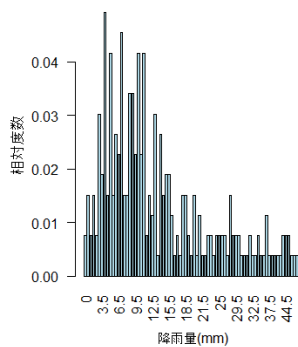


図 1 博多での相対度数分布

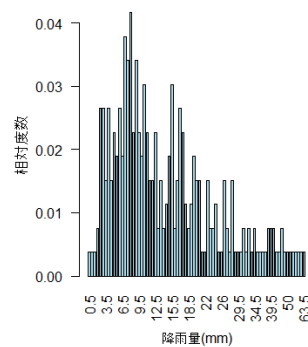


図 2 神戸での相対度数分布

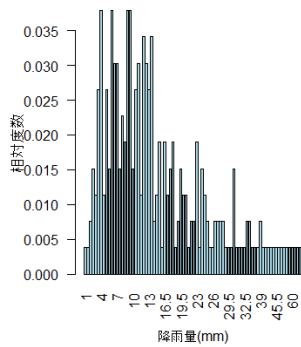


図3 名古屋での相対度数分布

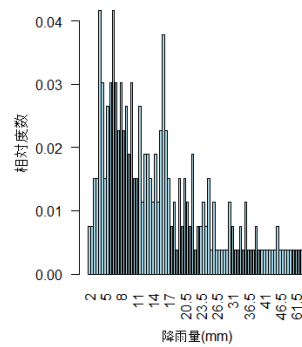


図4 さいたまでの相対度数分布

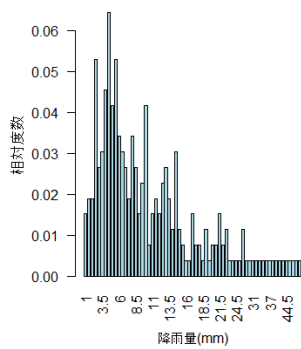


図5 仙台での相対度数分布

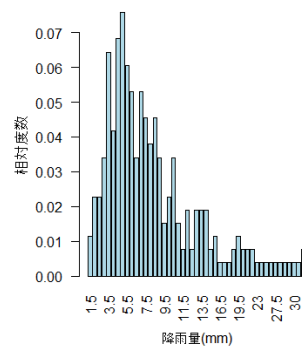


図6 札幌での相対度数分布

図1 から図6 に示すように、どの地点でも月最大降雨量の確率分布では、対称性が失われ、平均値より小さいところに最頻値を持つ非対称分布になることが分かる。

3 確率密度関数

3.1 採用した確率密度関数

ここで扱っているように、極値（まれにしか生じない値）の確率密度関数の形は様々なものが提案されている。正規分布や、対数変換によって正規分布に近似する方法が良く用いられている。従って、ここでは対数変換を基にしていく。

ここで採用した確率分布は、正規・対数正規・改良型対数正規・3母数対数正規及び改良型3母数対数正規の5種類である。また、正規分布以外は何いずれの分布型も、降雨を対数正規し、正規分布型に持ち込んで活用する。

$f(x)$ を確率密度関数とすると、確率指標 t は次式で定義される $F(x)$ の逆関数値で表される。

$$\int_{-\infty}^x f(t)dt$$

または

$$\int_x^{\infty} f(t)dt \quad (1)$$

ただし、 $f(t)$ が正規分布の場合は、

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2}\right) \quad (2)$$

となる。従って、 t の値としてどのような関数を使用するかによって、上述のような各種のパターンを表現できる。それをまとめると表2のようになるが、実数値 x の変形という意味で(2)式の t とは別の X を用いている。

表1 正規分布の指標 X の関数表示

形式	名称	X の関数型	備考
1	正規分布	$X = \frac{x_i - \mu}{\sigma}$	$\mu: x_i$ の平均, $\sigma: x_i$ の標準偏差
2	対数正規分布	$X = \frac{\log x_i - \mu}{\sigma}$	$\mu: \log(x_i)$ の平均, $\sigma: \log(x_i)$ の標準偏差
3	改良型対数正規分布	$X = \left(\frac{\log x_i - \mu}{\sigma}\right)^n$	$\mu: \log(x_i)$ の平均, $\sigma: \log(x_i)$ の標準偏差
4	3母数対数正規分布	$X = \frac{\log(x_i + b) - \mu}{\sigma}$	$\mu: \log(x_i + b)$ の平均, $\sigma: \log(x_i + b)$ の標準偏差
5	改良型3母数対数正規分布	$X = \left(\frac{\log(x_i + b) - \mu}{\sigma}\right)^n$	$\mu: \log(x_i + b)$ の平均, $\sigma: \log(x_i + b)$ の標準偏差

一般的に、パラメータの数を増加させれば、式の適合度が良くなることは当然である。そこで、できるだけ少ないパラメータ、精度よく確率密度関数が表現できる関数型が良い関数型ということになる。

確率密度関数が非対称であることから、少なくとも、2型は採用せざるを得ない。適合度を高めるためのパラメーターは3型は n 、4型では、 b の各1個であり、パラメータの数という点では同等の関数型と言える。

また、各関数型の適合度は次のように予測値 (計算値) h_i と実測値 x_i からの変換値 X_i の差の 2 乗平均誤差 (3) 式によって評価することにする。2 乗平均誤差を用いる理由は、小さな誤差よりも大きな誤差を重視するためである。

$$\epsilon_r = \sqrt{\frac{1}{N} \sum_{i=1}^N \epsilon_i^2} \quad (3)$$

i : 降水の観測値の順位

N : 降水の観測月数

ただし、 ϵ_i は次のように定義する。

$$\epsilon_i = h_i - X_i \quad (4)$$

h_i : 正規分布の逆関数値

X_i : ヘージンプロットの逆関数値

また、ヘージンプロットのプロットングポジション $P(x_i)$ は次の (5) 式によって定めた。ヘージンプロットを採用した理由は、観測データ数と順位さえ分かれば容易に非超過確率を決定できるためである。

$$P(x_i) = \frac{2i - 1}{2N} \quad (5)$$

記号は (3) 式と同じである。また、表 1 に示した指標 X に含まれるパラメータ n, b は (3) 式の ϵ_r が最小になるように定めた。ただし、対数値が負となる場合は絶対値をとり、マイナスをつけた。

3.2 適合性の検討

3.2.1 月最大降雨量の分析

博多・神戸・名古屋・さいたま・仙台・札幌の各地点における月雨量の適合度を (3) 式によって評価した結果表 2 のようになった。ただし、2 乗平均誤差を求めるにあたって、3 型では Nelder-Mead 法、4 型と 5 型では多変数 Newton-Raphson 法を用いてパラメータの最適値を求めた。

表 2 2 乗平均誤差 (ϵ_r)

形式	1	2	3	4	5
博多	5.502	1.573	0.985	1.160	0.985
神戸	4.693	3.013	1.106	1.260	1.089
名古屋	5.425	1.106	0.567	0.664	0.520
さいたま	4.342	4.167	1.245	1.431	1.253
仙台	4.704	1.916	0.730	1.432	0.730
札幌	3.270	1.248	0.654	0.655	0.655

表 3 パラメーター一覧

形式	3 型 (n)	4 型 (b)	5 型 (b)	5 型 (n)
博多	0.869	3.173	0.043	1.149
神戸	0.823	5.049	1.060	1.177
名古屋	0.928	1.861	0.723	1.053
さいたま	0.778	6.369	0.654	1.253
仙台	0.918	6.369	-0.167	1.099
札幌	1.071	-1.089	-1.093	1.000

これによると、1 型から 5 型にかけて二乗平均誤差が小さくなっていることから、パラメータの数が増えれば適合度が良くなるということが確認できた。また、パラメータの数と同じである 3 型と 4 型では、定数 b を加えた 4 型よりも、指数 n を付加した 3 型の方が適合性がやや良いようであった。従って、4 型よりも 3 型の方が良いことになる。

3.2.2 χ^2 検定とは

次に、 χ^2 法によって、2 型から 5 型の理論頻度分布と実測頻度分布の誤差を求めていく。これにあたって、 χ^2 検定について簡単に説明しておく。

χ^2 検定は、観測されたデータが期待されるデータの間からどの程度ずれているかを評価するための統計的手法である。 χ^2 値は次の式によって計算される。

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (6)$$

ただし、 O_i は観測値、 E_i は期待値である。 χ^2 値が大きいほど、観測値と期待値の間に大きな差があることを示し、小さいほど理論的に予測された値に近いことを示す。

3.2.3 χ^2 検定による評価

いかなの表 4 は、各地点における 2 型から 5 型までの χ^2 検定による結果を示したものである。

表 4 χ^2 検定における上側確率と適合性一覧表

	1 型	2 型	3 型	4 型	5 型
博多	0.000	0.266	0.126	0.126	0.171
神戸	0.000	0.211	0.112	0.112	0.032
名古屋	0.000	0.584	0.324	0.324	0.108
さいたま	0.000	0.326	0.064	0.637	0.438
仙台	0.000	0.346	0.147	0.147	0.189
札幌	0.000	0.264	0.291	0.291	0.250

有意水準を 5% としたとき、1 型は全ての地点で棄却されるため、適合性が極めて低いことが分かる。これは、月最大降雨量の分布が非対称であるため、対称な正規分布ではモデル化が困難なためであると推測できる。一方、2 型から 5 型は、いずれも χ^2 検定において有意水準 5% を超えているため、適合性が高いことが分かる。特に、2 型・3 型・4 型の適合度が最も高い結果となった。

3.2.4 同定期間と予測期間に分けた評価

これまで検討してきたのは、過去の観測値に最もよく合致する関数、もしくはできるだけ少ないパラメータで過去の観測値によく合致する関数型を抽出することであった。しかし、実際には将来の降雨量を予測することが重要であるため、過去の観測値に対する適合性だけでなく、将来の降雨量を予測する能力も評価する必要がある。そこで、観測資料を同定期間と予測期間に分けて、2 乗平均誤差による評価を行った。同定期間は 2003 年 1 月から 2018 年 12 月までの 16 年間、予測期間は 2019 年 1 月から 2024 年 3 月までの 5 年間とした。その結果を表 5、表 6 に示す。

表 5 同定時の誤差

	1 型	2 型	3 型	4 型	5 型
博多	0.460	0.080	0.076	0.072	0.072
神戸	0.348	0.120	0.116	0.120	0.120
名古屋	0.469	0.069	0.055	0.069	0.060
さいたま	0.385	0.192	0.137	0.192	0.094
仙台	0.440	0.100	0.100	0.100	0.093
札幌	0.522	0.113	0.096	0.113	0.109

表 6 予測時の誤差

	1 型	2 型	3 型	4 型	5 型
博多	0.367	0.115	0.108	0.121	0.121
神戸	0.545	0.198	0.185	0.198	0.068
名古屋	0.367	0.119	0.122	0.119	0.102
さいたま	0.412	0.240	0.233	0.240	0.210
仙台	0.561	0.203	0.203	0.203	0.200
札幌	0.494	0.144	0.133	0.144	0.132

この結果によると、同定時でも予測時でも 1 型は他の関数型に比べて大きな誤差を示している。これは、先ほどもあったように月最大降雨量の分布が非対称であるため、対称な正規分布ではモデル化が困難なためであると推測できる。また、同定時では 2 型から 5 型にかけて誤差が小さくなる傾向がある。しかし、予測時では 2 型から 5 型までの誤差が非常に近くなることが分かる。これは、パラメータの個数が異なっているにもかかわらず、予測の視点からは同じ精度になる場合があることを示している。つまり、複雑かつパラメータの多いモデルが必ずしも将来の予測において単純なモデルよりも優れているわけではないと推測できる。

4 確率密度関数の統合化

4.1 降雨の基準化

任意の集計時間に対する降雨量 x_i を次の式によって基準化し、統一的に扱いやすいように変数変換した。 x_i を任意の集計時間に対するある年の最大降雨量とし、 x_m をその平均降雨量とする。

$$X_i = \log x_i - \log x_m \quad (7)$$

x_i : 任意の集計時間に対するある年の最大降雨量

x_m : m 時系列の平均降雨量

4.2 数正規分布による確率雨量の統合化

第 3 章で分析したように、単純な対数正規分布でも、かなり良い適合性が得られた。この分布はパラメータを含まず、これによって、統合化ができる可能性がある。

統合化する際に最小二乗法をもちいるのでここで簡単に説明しておく。

4.3 最小二乗法

最小二乗法は、観測データに最もよく当てはまる回帰直線を求めるための方法である。回帰直線を $y = ax + b$ とするとき、 a と b を次の式によって求める。

$$a = \frac{S_{xy}}{S_x^2}, \quad b = \bar{y} - a\bar{x} \quad (8)$$

S_{xy} : 共分散

S_x^2 : 分散

\bar{x} : x の平均値

\bar{y} : y の平均値

4.4 統一化の結果

これから、観測場所に対する統合を考える。これまで得られた各 6 地点の直線を重ね合わせたのが図 7 である。さらに、この全観測値を用いて、最小二乗法により、直線の型をはめたのが図 7、平均である。この直線は確率指標 Y との間に次の関係がある。

$$Y = 0.952Z + 3.47 \times 10^{-15} \quad (9)$$

ただし、

$$Z = \frac{\log x_i - \log x_m}{\sigma}$$

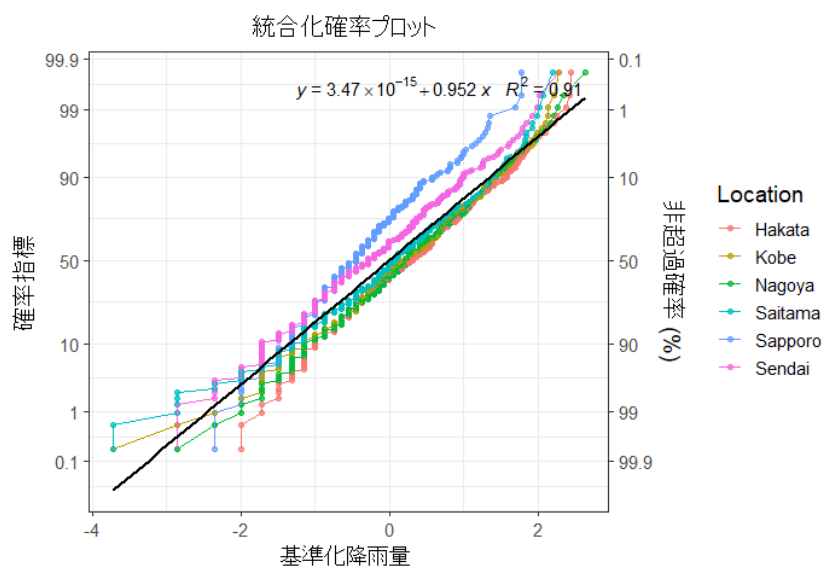


図 7 月最大降雨量統合化の確率分布

このように、観測場所に関わらず、一つの直線に統合化できた。改良型対数正規分布 (3 型), 3 母数対数正規分布 (4 型), 及び改良型 3 母数対数正規分布 (5 型) を適用すれば, さらに制度のよい統合化が可能であるが, b や n のパラメータが必要となり, 実用性に欠ける。

5 終わりに

今回は, 博多, 神戸, 名古屋, さいたま, 仙台, 札幌の 6 地点の月最大降雨量の頻度分布を用いて, その確率密度関数の検討を行った。その要点は以下の通りである。

- 月最大降雨量の確率分布は, いずれの地点でも非対称分布となり, 平均値よりも小さい箇所に最頻値をもちことから, 対称な正規分布ではモデル化が困難であることが分かった。
- 二乗平均誤差による評価では, パラメータの数が増えるに従って, 適合度が向上し, パラメータの数が同じである 3 型と 4 型では, 定数 b を加えた 4 型よりも, 指数 n を付加した 3 型の方がやや良いという結果が得られた。
- 優位水準 5% における χ^2 検定では, 1 型は全ての地点で棄却され, 適合性が極めて低いことが分かった。一方, 2 型から 5 型は, 全ての地点で有意水準 5% を超えており適合性が高いと判断され, 特に 2・3・4 型が最も高い適合度を示した。
- 同定時では 2 型から 5 型にかけて誤差が小さくなる傾向が見られたが, 予測時では 2 型から 5 型までの誤差が非常に近くなった。このことは, 予測の視点からはパラメータの個数が異なっても同じ精度になる場合があるということと推測できる。
- パラメータを含まない単純な対称正規分布に基づく降雨量の基準化を用いることによって, 観測場所に関わらず, 月最大降雨量の確率分布を一つの直線に統合化することができた。

結論として, 今回の研究では, 月最大降雨量の確率密度関数として, 対数正規分布をもとにした基準化を採用することにより, 観測場所や時間の違いに関わらず適用可能な, 実用性の高い確率雨量の統合化に成功したといえる。今後は, 今回試行していない 3 から 5 型でも同様の統合化を試み, その有効性を検討していきたい。

参考文献

- [1] 日本大学生物資源科学部: 劉剛・石川重雄・長坂貞郎, 石川県農業短期大学: 丸山利輔, 水文・水資源学会誌