水文・水資源学会誌
J. Japan Soc. Hydrol. & Water Resour.
Vol. 9, No. 2 (1996) pp. 136-142

3 母数対数正規分布の検討と改良分布型の作成

—カリフォルニア年降雨量を事例として—

Examination of The Logarismic Normal Distribution with 3 Parameters and Formation of a Modified Distribution—Survey of the annual rainfall in California—

志村 博康*・白岩 隆己*・石川 重雄* (日本大学農獣医学部)

Hiroyasu SHIMURA・Takami SHIRAIWA・Shigeo ISHIKAWA

Faculty of Agri. & Veterinary, Nihon Univ.

The annual rainfall data at Sacramento, Oroville and Shasta in California were tested whether they obey Normal distribution, Logarismic Normal distribution (with 2 parameters) or the distribution with 3 parameters. The last distribution was better than the others in agreement. Then, the calculating method of b , one of the 3 parameters which is added to rainfall data in order to agree with the distribution, was examined to introduce a simple method for founding b in minimizing the square of errors of a distribution graph.

However, the Logarismic distribution with 3 parameters has the limit to the agreement, especially around maximum or minimum. Then, a modified Logarismic distribution was prepared and proposed. The new distribution was better than the above in agreement. And the estimation of the errors was proposed in a view point which the errors of the nearer distribution to the true distribution approached to Normal distribution.

Key words: Logarismic Normal Distribution with 3 Parameters, Determination of b , Estimation of Errors, Modified Distribution, Annual Rainfall in California

カリフォルニア州の3地点、サクラメント、オロヴィール、シャスタの年降雨量を対象にして、正規分布、対数正規分布、3母数対数正規分布への適合性を検討し、最後者がより適合性がよいことを示すと共に、その母数の一つである降雨量に加算する値 b の算出法を見直し、誤差の2乗総和を最小にする b の算出法およびその簡便法を作成した。しかし、その b を用いても、3母数対数正規分布ではグラフ全体の適合性が必ずしも満足いくものではなく、特に極値付近では適合性が悪いことを踏まえ、改良型の分布を作成した。提案した手法は3母数対数正規分布の正規化方程式を n 乗 (ただし $0 < n < 1$) するもので、従来の対数型分布の手法を継承するものと言ってよい。その結果は、完全なる適合を生むものでは決していないが、従来の対数型諸分布よりはよい適合を示した。なお、この誘導過程でグラフ全体の誤差の評価について検討し、真の分布に近いグラフの誤差は正規分布に接近するという視点から、新しい評価法を作成した。

キーワード：3母数対数正規分布、 b の決定法、誤差評価、改良分布型、カリフォルニア年降雨量

I. はじめに

1980年代末から1990年代初頭にかけて、カリフォルニアは異常渇水に見舞われた。その実態を調査し、年降雨量の確率分布等を検討しているうち、3母数対数正規分布式の中の一つのパラメータ (分布形を合わせるために、降雨量 x に加算する或る数値、即ち $Ln(x+b)$ の b) のこれまでの決定方法に、幾つか疑問が

生じた。そこで、あらためてその決定方法を見直し、妥当と考えられる方法、その簡便法を作成した。次いで誤差の評価についても考察し、あらたな評価法を考案した。その上で、3母数対数正規分布式は、特に極値近くが降雨確率値と合わないことに注目し、合わせるための改良型分布を作成した。

ここで、用いたデータは、カリフォルニア州のサクラメント、オロヴィールおよびシャスタの3地点

*日本大学農獣医学部 〒252藤沢市亀井野1866, 電話0466-81-6241

Faculty of Agri. & Veterinary, Nihon University, 1866 Kameino, Fujisawa, 252, Japan

の年降雨量である。サクラメントはカリフォルニア中部の半乾燥地域に属し、稲作を含めて農業の主要地域であり、水の大消費地である。またオロヴィール、シャスタはカリフォルニア北部の水源地域に属し、それぞれ巨大なダムが存在する。用いた年降雨量は、サクラメントでは1878～1990年間、内欠損4年、有効年数109年、オロヴィールでは1949～1980年間、内欠損6年、有効年数26年、シャスタでは1950～1990年間、内欠損5年、有効年数36年である。

観測降雨値の確率値を求める方法として、トーマスプロットを用いた。この方法は周知の如く理論的基礎があり、観測降雨値の大きさに従って階級区分をした場合、ある階級に属する降雨の確率値の平均値を表すから、多くの観測降雨値を扱う我々の場合には、かなり有効であると考えられる。

II. 正規分布との照合

対数型に入る前に、まず正規分布との照合を見ておこう。図-1の通りである。

図-1の縦軸は確率指標を表す。ここでいう確率指標とは、標準正規分布式を用いて、確率値（非超過確率値）から求めた確率変数値（以下では t と記す）である。確率値そのものを用いずに、標準正規分布式の逆関数値 t を用いるのは、市販の確率グラフ用紙と同じで、与えられた確率分布のグラフが直線となり、観測値（ここでは、以下に述べるように、トーマスプロット値の標準正規分布式による逆変換値）との対照が明瞭になるからであるが、加えて、後に述べるように、誤差の計算に際しても便利かつ明瞭であるからである。なお、この確率指標は本論文での指標であるが、本論文では、一貫してこの確率指標を縦軸に用いることにする。

横軸は年降雨量 x であるが、 t と x は次の正規化方程式で結ばれている。

$$t = (x - \mu) / \sigma \quad (1)$$

ここに、 t ：標準正規分布の確率変数

x ：年降雨量

μ ：年降雨量の平均値（標本平均）

σ ：年降雨量の標準偏差（ただし標本なので不偏分散のルートを用いる）

正規化方程式は図上において直線となる。

もし降雨の確率値が標準正規分布に従うならば、トーマスプロット値と正規化方程式が一致する。

しかし図-1のトーマスプロット値は必ずしも直線ではない。両極が直線より右に出る傾向があり、異なる分布型であることを示唆する。

確率値の誤差は、色々の考え方があり、どれが正しいか、簡単にはいえないが、ここでは、次の式によった。観測降雨の確率値の誤差は、与えられた分布型を前提として、

誤差＝トーマスプロットから求めた確率値

－ 与えられた分布型の確率値

とする。即ち与えられた分布型から見れば、誤差はこのようになるという考え方である。

さて、先述のように、図上では、トーマスプロットの確率値は、標準正規分布の逆関数で表示された。また図上の標準正規分布の確率値は(1)式で表示された。よって、図上での誤差は、それぞれの確率変数について、(1)の直線から、トーマスプロットの標準正規分布式による逆変換値が、どれだけ離れているか、を表示するものとなる。

グラフ全体としてのトータルな誤差は、上の値の二乗の総和で与えることにした。図中には、そのグラフの誤差の二乗総和の値が表示されている。

図-1 年降雨量と正規分布との比較

Ⅲ. 対数正規分布との照合

図-2の通りである。

縦軸は先の図-1と同じである。横軸は年降雨量の対数となる。 x から t への変数変換が変わり、正規化方程式は次のように変わる。

$$t = (Ln x - \mu) / \sigma, \quad Ln: \text{自然対数}$$

なお μ は $Ln x$ の平均値であり、 σ は $Ln x$ の標準偏差となる。

年降雨量が、もし対数正規分布に従うとすれば、図中のデータは正規化方程式と一致しなければならない。サクラメントはかなり合うが、オロヴィール、シャスタは、極値付近のデータは直線から遠ざかり、別の分布型を求めているといってよい。

誤差の二乗総和を求めると、正規分布よりは減っているが、なおかなりの誤差をもつ。サクラメントも、図上ではほぼ照合しているように見えるが、やはりかなりの誤差をもつ。

Ⅳ. 3母数対数正規分布との照合

図-3の通りである。縦軸は前と同じ。横軸は、年降雨量に或る数値 b を加算したものの対数となる。 b があらたに求めるパラメータである。

変数変換が変わり、

$$t = (Ln(x+b) - \mu) / \sigma$$

b : ある数値 (母数の一つ)

となる。ここに μ は $Ln(x+b)$ の平均値、 σ はその標準偏差となる。

1. b の決定

日本では、岩井法 (岩井・石黒, 1970) が簡便なので、よく用いられる。しかし後に見るように、広い概算値を与えるものであって、正確とは必ずしもいえない。石原・高瀬法もよく知られているが、複雑で、簡便とはいえない。

けれども、 b の値が何を基準にして正確というのか、必ずしも明確ではない。ここでは、分布型から

図-2 年降雨量と対数正規分布との比較

図-3 年降雨量と3母数対数正規分布との比較

見て、トーマスプロットの逆関数値がどれだけ離れているか、を見るという立場から、上記の誤差の二乗総和を判定に用い、それを最小にする b を最適解とした。

1) b の計算と簡便法の大枠

誤差の二乗総和を最小にするを求めるには、通常の方法としては、誤差の二乗総和を表示する式を b で微分し、それを0とおいて、 b を得る。だが、我々の場合には、 b で微分した式は一層複雑となり、その上、 b を陽に解く形式にならないため、むしろ誤差の二乗総和を表示する式そのものに、適当な b を幾つか与え、それを最小にする b を決めるのが無難である。以下ではその方法を用いる。

パソコンの普及した今日、誤差の二乗総和を求め、その最小値を与える b を求めることは、比較的容易となったが、簡便法があれば、便利なことはいうまでもない。

誤差は極値に近いところが大きいという傾向がある。そこで、極値から順に二乗誤差を求める。ただし、最大極値、最小極値、最大から2番目の極値、最小から2番目の極値という様に大きいものから、次いで小さいものから、を交互に順序よく取り上げて、次第に中心部に及んでいく。求めた二乗誤差を、順次、加算していく。これを中心部まで計算を行えば、結局、二乗誤差の総和を求めたことになる。

それぞれの二乗誤差を求める時、 b を適当に与えることが必要であるが、 b を、例えば、-100, 0, 100, 200, と入れて二乗誤差を計算する。すると、計算の途中において、二乗誤差を最小にする b が次第に鮮明になってくる。極値に誤差が集中している場合には、全体の約1割程度の極値に近い部分を計算するだけで、 b がほぼ推定できる。誤差が全体にほぼ一様にばらまかれているような場合には、極値からほぼ中心部まで計算しなければ b が定まらないから、結局、二乗総和誤差を計算したことになる。しかし、この計算過程で、誤差がどのように分布しているかがわかるので、計算が無駄であったということにはならないであろう。

以上を数式化しよう。まず、図中の各地点の記号化から始める。最大極値を $L1$ 、最大から2番目の極値を $L2$ 、同3番目を $L3$ (以下同様)と名付ける。最小極値を $S1$ 、最小から2番目の極値を $S2$ 、同3番目を $S3$ (以下同様)と名付ける。

次に、トーマスプロットの逆関数値を Th と書き、

以上の各地点の Th 値を、上記の順に従って Th_{L1} , Th_{L2} , Th_{L3} , Th_{S1} , Th_{S2} , Th_{S3} (以下同様)と書く。次に、正規化方程式の値を St と書き、以上の各地点の $Ln(x+b)$ が与える St 値を、上記順に従って、 St_{L1} , St_{L2} , St_{L3} , St_{S1} , St_{S2} , St_{S3} (以下同様)と書く。

これらを順に並べると以下ようになる。

$L1$	$L2$	$L3$	$S1$	$S2$	$S3$
Th_{L1}	Th_{L2}	Th_{L3}	Th_{S1}	Th_{S2}	Th_{S3}
St_{L1}	St_{L2}	St_{L3}	St_{S1}	St_{S2}	St_{S3}

ここで留意しておくべきことは、 Th は b が変化しても、不動であるが、 St は変化することである。

誤差の二乗総和 E は次式で与えられる。

$$E = (Th_{L1} - St_{L1})^2 + (Th_{S1} - St_{S1})^2 \\ + (Th_{L2} - St_{L2})^2 + (Th_{S2} - St_{S2})^2 \\ + (Th_{L3} - St_{L3})^2 + (Th_{S3} - St_{S3})^2 \\ + \dots$$

この E 式は、事前に与えられた b_j の数だけ存在するが、 E を最小にする b_j が計算の途中から次第にわかってくる。

2) b の概算に当たっての μ および σ の略算

以上が b の概算法の大枠であるが、その計算を進めるに当たっては、なお苦労が残っている。それは St が b によって変化するが、 μ および σ も b によって変化するの、 b が変化する度毎に μ および σ を計算しなければならないことである。これはパソコンを用いても少し面倒なことである。そこで、次の略算法を案出した。

[μ の概算]

正規化方程式を表す図中の直線が b によってどのように変化するかを見ると、 b が増大する場合において、その直線は、次第に勾配を増大しつつ、右方向(横軸の値の増大方向)に移動していく。直線の中心点を表す μ は、概算として、横軸の値の増加に合わせて増大すると判断される。実際、図上で測ってみると、ほぼその通りであった。

$$\mu = Ln(x_0 + b)$$

x_0 : 任意の計算開始点

ただし、この式が有効な範囲は、我々の計算の範囲(降雨値 x が数100mm~2,000mm程度)では、 b を動かし得る範囲が0~500程度以内であった。 b をこれ以上の範囲に広げる場合には、よい結果が得られなかった。

しかし、上記の範囲で間隔を選んで、計算開始点

を数点選定し、計算を進めれば問題はない。

【 σ の概算】

先述のように b の増大に対して正規化方程式の直線は勾配を増加させるが、これは、正規化方程式および σ を求める不変分散式から理論的に判断して、 μ の増大および σ の減少の結果である。すなわち σ は b の増加に対して減少する。その減少の関係式は、不変分散式から判断して、 σ の二乗にて b と関係すると考えられる。そこで

$$\sigma^2 = \sigma_0^2 - \alpha b$$

と置く。ここに σ_0 は任意の計算開始点の σ 、 α はある常数で、 b を動かす区間の両側の σ (観測データより算出) から、前もって求めておくものである。

我々の計算の範囲では、 b が0～300程度以内で有効であったが、それ以上に範囲を広げると、よい結果が得られなかった。

μ および σ の略算に有効な b の範囲は、降雨量の大きさに関係しているようで、 b が降雨の値より十分に小さい時には有効であるが、降雨が小さくなると、合わなくなるといってよい。略算式を用いる場合には、 b の計算範囲を降雨の値と見比べて、大きくとらないことが大切である。もっとも、 b そのものの値が降雨に比べて著しく大きくなるようでは、少し問題があるといえよう。データ処理の問題なのか、あるいは3母数対数正規分布に合わない分布なのか等も検討する必要がある。

サクラメント、オロヴィール、シャスタのそれぞれについて、 b の変化による E の変化を求めたのが、表一1である。岩井法で求めた結果も挿入した。

b による E の変化は大きいものでないことがわかる。誤差を減らすために b に期待をかけても、思った

表一1 b による誤差の二乗総和の変化と岩井法の結果

b	-650	-100	0	100	150	190	200	300	1000	1170
サクラメント			1.016	0.485	0.435	0.4236	0.4246	0.539		
オロヴィール		0.913	0.803	0.769	0.767		0.771		1.004	1.051
シャスタ	1.993	1.849	1.785	1.755			1.748	1.755		

— は E の最小値を表示。

岩井法の計算結果

	総データ数	計算データ数	岩井法の b
サクラメント	109	22	108
オロヴィール	26	6	1173
シャスタ	36	8	-643

計算データは最大値と最小値を対にして順に取り出す。

程には変化しないことに注意すべきである。

岩井法の結果は E の最小値からかなり離れることがある。岩井法が意外な値を出すことがあるのは、簡単に述べると、同法は、対とする大きい側と小さい側のそれぞれの点と中心点が一直線になることを前提にして、組立てられているが、 b が必ずその直線を描くとは限らないからである。

2. 誤差の評価

誤差の二乗総和が、各分布型への一致性について、よい判定になることは言うまでもないが、誤差の大きさが必ずしもその分布型でよいと保証するものではない。そこで、二乗総和以外の誤差評価を考えてみたい。

一般的にいつて、誤差の生起条件が総体としてほぼ同一の場合、その偶然誤差は、充分なデータを前提として、正規分布になりやすいとみられている。我々の領域に広げれば、年降雨量の生起条件が、充分な期間をとって、ほぼ同一とみなされる場合(すなわち、一つの分布型でほぼ表示される場合)、その誤差(その分布型からのはずれ)は、正規分布に近くなると考えられる。このことは、正規分布が、充分なデータに基づく場合、統計の主要な分布関数の母体あるいは極限になり、時系列における不確定性原理の最小条件となること等を勘案すると、無理からぬことのように思われる。

すると、もしデータ(トーマスプロットの逆関数値)がある分布型によく合致する、あるいはその分布型こそ総合的にみて生起条件をほぼ同一化しているという場合には、その分布型の誤差は正規誤差分布に近いであろう。ここに言う誤差は確率変数のそれぞれに対応する確率値(本論文ではトーマスプロットの逆関数値)の誤差で、それらが全体として正規誤差分布になるという見方である。逆に、その分布型に合わないものであれば、その誤差は系統誤差を含んでおり、別の傾向、トレンドをもつと考えられる(ここで正規誤差分布と述べたが、正規分布と同一である。先に見てきた観測値の正規分布への照合とは全く異なる議論であるので、誤解のないように正規誤差分布といったままである)。

以上の考え方に基いて、3母数対数正規分布の誤差が正規誤差分布に従うか、否か、を見たのが図一4である。ここで、誤差の生起確率はトーマスプロットで与えた。

誤差の分布は中心部分で、ほぼ、正規誤差分布に

図一 4 3 母数対数正規分布に対する誤差と正規誤差分布との比較

従うようであるが、極値付近では、やはり、正規誤差分布とは異なる分布型を求めているように感じられる。サクラメントは3母数対数正規分布でほぼよいように思われたが、なお誤差があるし、オロヴィール、シャスタでは、別の分布型が望まれるであろう。

ここでは割愛したが、正規分布、対数正規分布の場合にも、それらの誤差の分布を正規誤差分布と比較対照したが、いずれも合わず、別の分布、トレンドの存在を想起させるものであった。

もっとも、観測値自体が間違っており、従ってトーマスプロットの逆関数値が違っているということも考えられる。即ち極値付近の直線から著しく離れた点は異常値と考えた方がよい場合もある。しかしここでは、信用の優先度をまず観測値に置き、できるだけ分布型を接近させることにした。極値付近の降雨量を分布型から推測する場合、極値付近の観測降雨値を異常値として排除すれば、豪雨では極めて大きい値、渇水では極めて小さい値を出してしまうであろう。極値付近でも観測値を基にして分布型をきめるのが無難であるように思われる。

V. 3 母数対数正規分布の改良

新たな分布型では、対数型を継承することが望まれるであろう。これまでの分布型の採用には、次第に改良を重ね、真の分布型に近づくとする流れがあるからである。そこで、これまで検討を重ねてきた対数型を基礎とし、その改良型を考えることにした。

標準正規分布の確率変数 t の変数変換を次のように置くことにする。

$$t = ((\ln(x+b) - \mu) / \sigma)^n$$

n は任意に取れるが、図上で、データ(トーマスプロットの逆関数値)が正の場合には直線(正規化方程式)の下側、負の場合には直線の上側になる傾向があることを考慮すると、 n は、 $1 > n > 0$ の値を取るものと考えられる。しかし

$$((\ln(x+b) - \mu) / \sigma) < 0$$

の場合には、 t の値が定まらないので、その絶対値を取って計算を行い、その後には $-$ の符号を付けることにした。

$$t = (-) | (\ln(x+b) - \mu) / \sigma |^n$$

ここに $(-)$ は、絶対値の中が負の場合に付けることを意味する。

この改良法で計算した結果が図一5である。サクラメントは3母数対数正規分布にほぼ近かったので、 n の値は1に近く、0.95であったが、オロヴィールでは1から離れて0.70、シャスタでは0.77となった。

サクラメント、オロヴィール、シャスタいずれもほぼ良好な結果を示している。参考までに誤差の二乗総和を示したが、3母数対数正規分布よりさらに小さい値となっている。

改良型の誤差を評価するために、その誤差が正規誤差分布に近いかどうか、を見たのが、図一6である。正規誤差分布によく一致しているとは必ずしもいえないが、3母数対数正規分布の誤差よりも、改良型の誤差の方がより正規誤差分布に近いように思われる。

改良型が一般的に有効かどうか。今のところ不明であるが、大雑把に見てカリフォルニアでは有効であるように思われる。これにより極値に近い年降雨量の推測に改善が図られるであろう。

なお、改良型分布の確率密度関数はここでは割愛するが、少し複雑な式となる。けれども標準正規分

図一五 年降雨量と改良分布型との比較

図一六 改良分布型に対する誤差と正規誤差分布との比較

布とつながられているので, 比較的扱いやすいと考えられる。

VI. 結 び

カリフォルニア州の3地点, サクラメント, オロヴィール, シャスタの年降雨量を取り上げ, 正規分布, 対数正規分布, 3母数対数正規分布の順に, それらの分布型に合うかどうか, を検討してきた。この順で次第に合うようになってきたが, もう一步, 合う方向をめざして新分布型に向かうことが望まれた。

この間, 3母数対数正規分布の一母数である $Ln(x+b)$ の b を決める方法が, これまでのものでは必ずしも満足出来ないで, 新たな決定法, その簡便法を作成した。また誤差の評価について, 異常値の検討方法はあっても, 誤差を全体として適正に検討する方法が必ずしも提示されていなかったで, 新し

い誤差評価の方法を作成した。

その上で, 対数型分布を乗り越える新しい改良型分布を作成した。この分布型が, 上述分布型と比べ, 誤差がもっとも少なかった。

なお, この改良型は, 標準正規分布の確率変数 t を前提にして, 対数型をそのまま取り入れながら変数変換をしているから, ある意味で対数型の延長上にあり, 対数型で経験したことが色々役立つように思われ, それなりの利便性があるように思われる。

カリフォルニア地域の水の状況把握, 水文データの収集等に際しては, 東京大学農業水利学研究室から多大の御支援を頂いた。心から御礼申し上げる。

参考文献

岩井重久・石黒政儀 (1970) : 応用水文統計学, pp76~77

(1995年10月23日受付, 1995年12月4日受理)