# 対数正規分布の活用

～正規分布の限界と対数正規分布との使い分けに関する考察～

近藤恭彦

会津大学短期大学部研究紀要　第 79 号抜刷
2022 年 3 月

# 対数正規分布の活用

## ～正規分布の限界と対数正規分布との使い分けに関する考察～

近藤恭彦*

【要旨】現在、統計処理において、正規分布は非常に多用されている。観測値が正規分布に従うと想定される場合が極めて多いからである。然るに、正規分布は原則的に、データの値が-∞から+∞までの値を取り得る場合でなければ成り立たない。だから何らかの長さや質量など、正の値しか取り得ない「大きさ」の統計データは厳密には正規分布に従う事はない。

正の値しか取らない統計データでも、各データ $x$ の対数の値をとると、その値は負の値を取り得 る。即ち、$0< x$ のとき、$-\infty < \log x < +\infty$ となる。そして、その対数の分布が正規分布に従う分布は色々存在する。このように正の値しかとらないデータの値の対数をとると正規分布に従う分布になるものを対数正規分布という。自然現象の「大きさ」は倍数の世界（指数関数的、 等比数列的）であるので、対数正規分布近似は有効な確率モデルであると考え得る。

正の値しか取り得ないデータでも、かなり有効な正規分布近似が出来る分布も存在する。それにはどのような条件が必要であるかをグラフを描いて調べてみると、平均値$\mu$に対して標準偏差$\sigma$が十分に小さいこと: $\sigma \ll \mu$ が条件であった。そこから、人の身長は正規分布でかなりきれいに近似出来るが、体重の正規分布近似は精度が良くないという面白い考察が得られた。

負の値を取り得る統計データには、正規分布を使うべきであるが、何らかの物理量の大きさのように、正の値しか取り得ない統計データの分布は、原則的に対数正規分布で近似したほうが良いと考えられる。従来、正の値しか取らない統計データの多くが、正規分布に近似してデータ処理が行われてきた。しかし実は対数正規分布に近似させたほうが、実際のデータの分布をより適切に表現している場合が多いと考えられる。統計処理に当たっては、正規分布近似と対数正規分布近似とを比較して、実際のデータの分布をより正確に記述出来るほうを使うべきである。

* 会津大学短期大学部非常勤講師

## １．序

数理統計学では、観測された統計データが、何らかの確率法則性に従う現象の1つの実現値であると見なす事によって資料を分析する。このとき想定される確率的法則性は確率モデルと呼ばれている。データ分析にあたっては、数理統計学を有効に適用できていれば、想定された確率モデルが現実を適切に表現できる。

現在、正規分布は、その名の示す通り、非常に多くの事象の確率モデルとして選ばれている。その際、特に検討せずに、はじめから正規分布に従うものとして正規分布近似している場合が大部分であるようにみえる。その中には疑問を感じるものも多い。正規分布近似が相応しくないようなデータ解析にも正規分布が堂々と用いられていると思われるケースもある。正規分布近似よりも寧ろ対数正規分布近似のほうが相応しいと考えられる統計データはかなり多いと考えられる。そのことについて論述し、正規分布と対数正規分布の特徴と使い分けに関して考察する。

## ２．相加平均と相乗平均

確率モデルとして、正規分布と対数正規分布を対比する前に、関連事項として相加平均と相乗平均の代表値としての優位性について考察する。

n個の正数 $x_i (i=1,\ 2,\ 3,\ \cdots,\ n)$ の 相加平均を $\overline{x_a}$ 、相乗平均を $\overline{x_g}$ とすると、

相加平均　$\overline{x_a}=\dfrac{x_1+x_2+x_3+\cdots+x_n}{n}$

相乗平均　$\overline{x_g}=\sqrt[n]{x_1\cdot x_1\cdot x_3\cdot\cdots\cdot x_n}\quad=(x_1\cdot x_2\cdot x_3\cdot\cdots\cdot x_n)^{\frac{1}{n}}$

である。日常生活では一般に「平均」と言えば相加平均しか使わない。相乗平均は高校で習った、
（相加平均）≧（相乗平均） の関係でしか使わない、使えない平均であると思っている人も多いと思われる。然るに実は世の中の現象、自然現象を表す代表値としての平均値は、相加平均よりも相乗平均のほうが相応しい場合が多い。その例を示す前に、先ずは相乗平均とはどのような平均値であるかをイメージする。

以下、イメージし易いように、データ数2個の正数の平均について述べる。

2正数の相加平均:　$\overline{x_a}=\dfrac{x_1+x_2}{2}$　、 相乗平均:　$x_g=\sqrt{x_1\cdot x_x}=(x_1\cdot x_2)^{\frac{1}{2}}$　である。

具体的な2つの数値の例で相加平均と相乗平均を比べてみる。

①．$x_1=2,\ \ x_2=8$　の場合

相加平均: $\overline{x_a}=\dfrac{2+8}{2}=5$　相乗平均: $\overline{x_g}=\sqrt{2\cdot 8}=4$

相加平均5は、小さいほうの値2より3大きく、大きいほうの値8より3小さい。

一方、相乗平均の4は、小さいほうの値2の2倍、大きいほうの値8の $\dfrac{1}{2}$ 倍である。

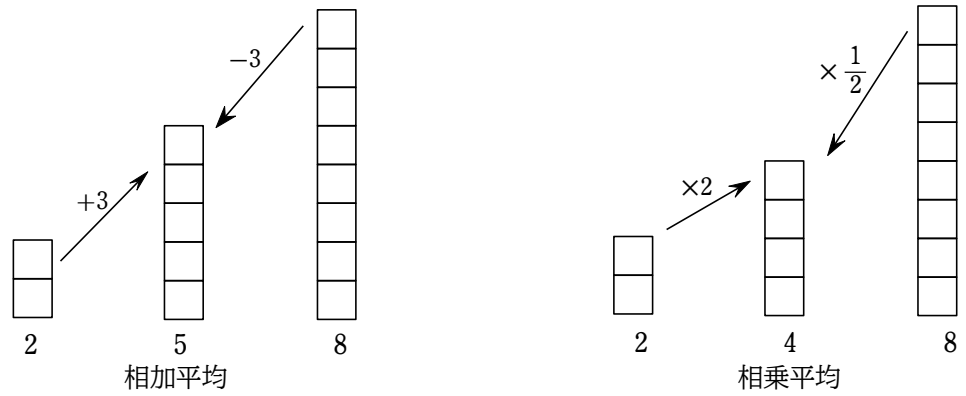


**図１－1.　2と8の相加平均と相乗平均のイメージ**

②. $x_1=1,\ x_2=9$　の場合

相加平均: $\overline{x_a}=\frac{1+9}{2}=5$　　相乗平均: $\overline{x_g}=\sqrt{1\cdot 9}=3$

相加平均5は、小さいほうの値1より4大きく、大きいほうの値9より4小さい。

一方、相乗平均の3は、小さいほうの値1の3倍、大きいほうの値9の $\frac{1}{3}$ 倍である。

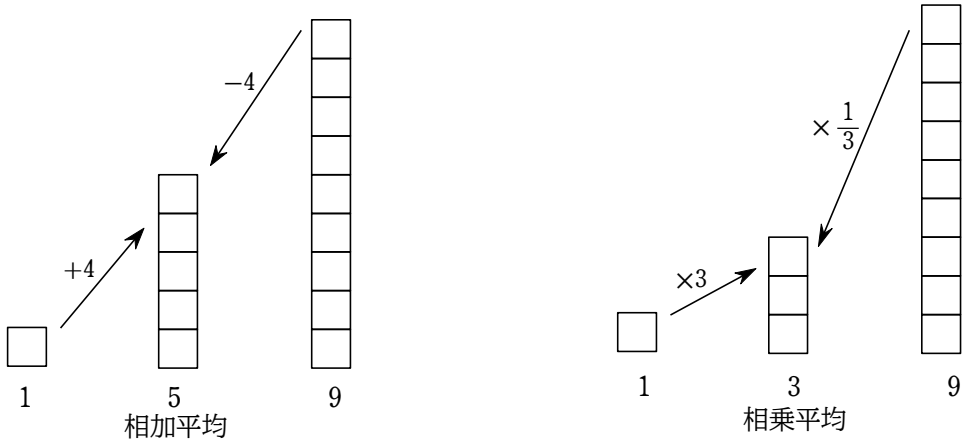


**図１－2.　1と9の相加平均と相乗平均のイメージ**

以上より

相加平均は、自然現象、世の中を等差数列的（線形）に捉えた場合の代表値としての平均値

相乗平均は、自然現象、世の中を等比数列的（指数関数的）に捉えた場合の代表値としての平均値

であると考えられる。

はたして、ほとんどの事象に於いて、相加平均のほうが適切な平均値と言えるのであろうか。相乗平均のほうが適切であると考えられる例を挙げてみる。

音名A4の音は、図2のように五線譜の真ん中の「ラ」の音で、振動数は440Hzである。

これよりも1オクターブ下の「ラ」は音名A3 で振動数は220Hz 、1オクターブ上の「ラ」は 音名A5 で振動数は880 Hzである。A3 と A5 の音の振動数の相乗平均は　$\sqrt{220\cdot 880}=440\,(\mathrm{Hz})$　となって丁度 A4の音の振動数に

なる。

ちなみにA3 と A5の音の振動数の相加平均は

$\frac{220+880}{2}=550$ (Hz)　となる。550HzはA4よりも高い音になる。これは 、五線譜の真ん中の「ラ」の音 A4 の少し上の「ド」の音を半音上げたドのシャープ（♯）の音 C5♯の音高とほぼ同じ高さ（わずかに低い）である。つまり、音の高さを振動数で表した場合、相加平均はほとんど無意味である。それに対し、相乗平均は的確な意味があることがわかる。

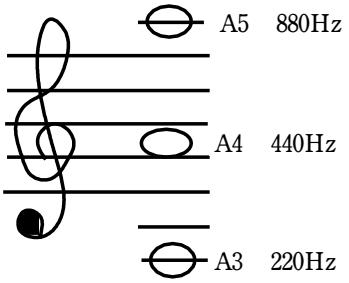


**図2.　A音の高さと振動数**

オクターブの例でなくとも、例えば並んだ「ド」と「レ」の音、「レ」と「ミ」の音は共に音程が全音である。どの高さでも隣り合った「ドレミ」の音の「ド」と「ミ」の振動数の相乗平均は「レ」の振動数になっている。しかし、「ド」と「ミ」の振動数の相加平均はその二つの間の「レ」よりもわずかに高い音になり、相加平均の音高としての意味は特に見あたらない。

音の高さだけではない。音の大きさも相乗平均のほうがしっかりした意味がある。音の大きさは、音圧、若しくは音の強さであらわす。単位はdB（デジベル）である。

音圧は20dBで10倍と決められている。音の強さは音圧の2乗に比例するので、20dBで $10^2=100$ 倍である。静かな事務所の中がおよそ50dB、騒々しい事務所の中がおよそ70dBといわれているで、この2ヶ所のdBの値の相加平均は $\frac{50+70}{2}=60$dB　である。この60dBは、dBという形式では、相加平均になっていて大いに意味がある。これを音圧、若しくは音の強さで考えると、60dBは、50dBと70dBの相乗平均になっている。50dBの音圧を1とすると、70dBの音圧は10で、60dBの音圧は、その相乗平均で $\sqrt{1\cdot 10}=\sqrt{10}$　となっている。音の強さで言えば、50dBの音の強さを1とすると、70dBの音の強さは $10^2=100$ で、60dBの音の強さは、その相乗平均で $\sqrt{1\cdot 100}=10$となっている。dBは音圧の対数表示で表しているのである。（対数の底は10）

以上、人間が2つの音の丁度中間ぐらいの「高さ」、若しくは「大きさ」だと感じる音は、それぞれ相乗平均になっている。

音の高さ・強さは、等差数列的に変化しても異様に感じ、人間（生物）の感覚にそぐわない。一方、等比数列的に変化すると、しっかり規則的に変化しているように感じるのである。

以上、音の高さ、強さに関する平均値としては、相乗平均が有意義で、相加平均はほとんど無意味である事を示した。自然現象は、桁数が極端に違う値を表記しなければならないので、一般に対数表示になっている単位が多い。その場合、形式的に相加平均として求めた値が、エネルギーや強度で考えれば相乗平均になっているのである。地震の震源のエネルギーの大きさであるマグニチュードでも、酸性度を表すpHでも、天体の明るさを表す等級（magnitude ）でも、対数の値を使っているので、その値の相加平均は、エネルギーや強度で記述すれば相乗平均になっているのである。ただし、酸性度：pHや天体の明るさ：等級　は逆対数スケール（対数の値の逆符号）として定義されているので、値が小さいほど大きくなる。

エネルギーや強度の平均値は相乗平均で表すほうが便利であるという事は、自然現象の変化を対数関数的に捉えたほうがいいという事でもある。自然界の色々な量の変化を人間（動物）が感知するためには、何百万倍、何十億倍、何兆倍のオーダー［ order of magnitude］の違いを感知しなければならないから、人間（動物）の感覚器も対数関数的に感知するようになっているのであろう。

つまり、音でも光でも、そのエネルギーの量は小さい値から桁違いに大きい値まで数万倍～数兆倍、くらいのオーダーを扱うので、その変化を感知する為には、人間の感覚器は、対数関数的に感じ取る感覚器のほうが好都合であり、単位としても対数表示的なものが適している。逆に言えば、リニアな(等差数列的な)変化は感知しにくく、指数関数的(等比数列的)な変化を人間(生物)は感知しやすいという事であろう。自然界の変化が指数関数的なのに対して、人間の感覚が対数関数的になっていて、その結果、自然界の指数関数的変化(等比数列的変化)を、人間の感覚はリニア的な変化(等差数列的変化)に感じると考えられる。

相加平均は等差数列的平均、相乗平均は等比数列的平均と考え得る。その相加平均、相乗平均の関係と同様の関係が、分布の確率モデルとしての、正規分布と対数正規分布の関係にあると考えられる。すなわち、分析する対象を相加平均のようにリニア(等差数列的)に扱うか、相乗平均のように、指数関数的(等比数列的)に扱うかという事である。

## ３． 正規分布近似の誤適用

連続分布の中で、最も広く、普通に用いられているのは正規分布[normal distribution]である。極めて有用な理論的確率分布と捉えられているが、本当に有用で適応範囲の広い理論的確率分布であるのであろうか。その欠点について考察する。

正規分布の確率密度関数 [p.d.f.: probability density function] $f(x)$ は

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\ -\infty< x <\infty\ )$$

$\mu$:平均値 、 $\sigma$:標準偏差 ( $e$ は自然対数の底 )

である。正規分布は一般に $N(\mu,\sigma^2)$ の記号で表す。具体的なグラフは図3－1、図3－2 のようになる。

(平均値$\mu$＝0 , 標準偏差$\sigma$＝1のときの正規分布を標準正規分布 $N(0,1^2)$ という:図3－1)

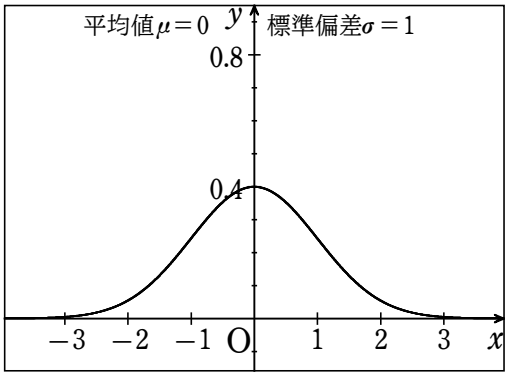


図３－１ 標準正規分布 $N(0,1^2)$

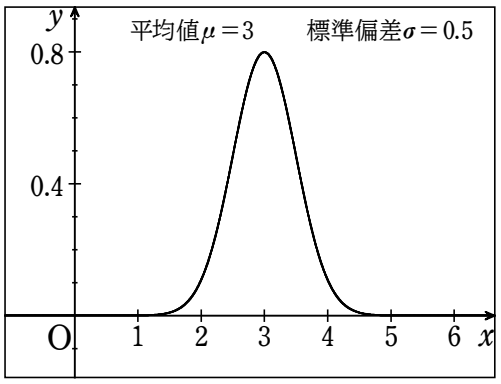


図３－２ 正規分布 $N(3,0.5^2)$

正規分布のグラフは、図3-1、図3-2 のように平均値$\mu$を中心にして左右対称である。だから、平均値と最頻値[mode:モード、ピーク]と中央値[median:メジアン]が一致する。

正規分布は、自然界・社会 に多く見られる分布に対する理論的確率分布モデルとして極めて有用で、大量の要素を持つ母集団の測定結果は、ほとんど正規分布によって近似されると言われている。

例えば

- 同じ年齢の人の身長や体重など、身体の様々な測定結果の分布
- 規格の決まった工業製品の(部品の)各部分の長さや質量の測定値の分布

など枚挙にいとまがない。…というよりも、現代社会では、適用できるかどうかの検討もせずに、何でもかんでも正規分布にあてはめようとする傾向がある。

中心極限定理（母集団の分布がどんな分布であっても、そこから任意に抽出された標本平均は、標本の数を十分に大きくしていくと正規分布に近づいていくという定理）により、独立な多数の因子の和として表される確率変数は確かに正規分布に従う。このことによって正規分布は統計学や自然科学、社会科学の様々な場面での複雑な現象を簡単に表すモデルとして用いられている。

正規分布はもともとカール フリードリヒ ガウス[ Carolus Fridericus Gauss]が測定誤差に関して詳細に論じたものであり、ガウス分布[ Gaussian distribution]とも呼ばれている。ガウスの誤差理論では、偶然誤差（データの収集方法が適切でないため系統的に起こる一定の方向性をもつ「系統誤差」に対して、偶然に起こる誤差を「偶然誤差」という）について経験にもとづく次のガウスの公理がある。

1.大きさの等しい正と負の誤差は等しい確率で生じる。

2.小さい誤差は大きい誤差より起こりやすい。

3.ある限界値より大きな誤差は実際上起こらない。

この、偶然の誤差についての3つのガウスの公理は、理性的に正しいと考えられる。ガウス分布＝正規分布は、ガウスの公理を満たす。工場における規格の決まった製品の長さ、質量などに関する測定値の誤差の分布は確かに正規分布に従うであろう。

では、現在一般に正規分布とみなされている他の事象の分布もガウスの公理を満たしていると考えて良いのであろうか。ガウスの公理を満たさない例を挙げてみる。例えば、同年齢の人の体重について考える。この場合ガウスの公理の「誤差」は「偏差（平均値との差）」と言い換えることとする。ガウスの公理の2番目、3番目は正しいと考えられても、1番目は正しいとは言えない。その例を以下に示す。

ある年齢の人々の体重の平均値を$\mu$とする。偏差の絶対値が 平均値$\mu$と同じ値の場合を考える。$\mu+\mu=2\mu$　、$\mu-\mu=0$ である。すなわち偏差が $+\mu$ の人の体重は、平均値$\mu$の2倍の $2\mu$になる。体重が$2\mu$以上の人はかなり少ないが、確かにある程度存在する。しかし逆に、偏差が $-\mu$ の人の体重は0になってしまうから、そういう人は存在しない。当然体重は正の値しか取らないので、体重に0以下の値は存在しない。体重が $2\mu$ 以上の人は存在し得るのに、体重が0以下の人が存在しないという事は、体重の分布が平均値を軸として左右対称にならないという事である。

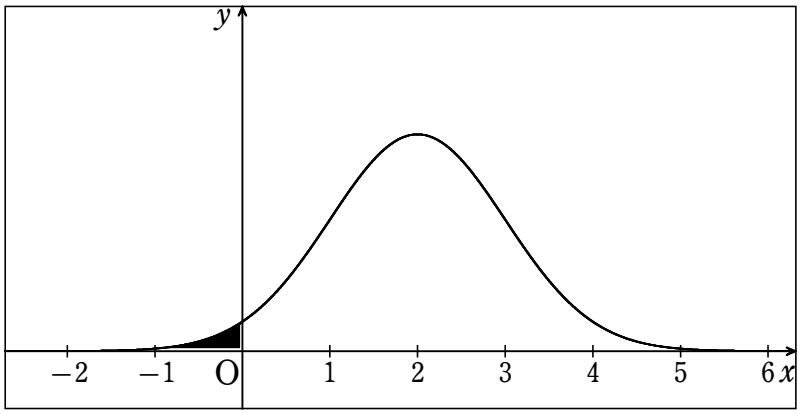


**図４.正規分布 の限界：定義域 $x>0$ のとき、矛盾が生じる。**

だから

1.大きさの等しい正と負の偏差は等しい確率で生じる。

は正しいとは言えない。図4のように、$x \leqq 0$ の部分(黒塗りつぶしの部分)に確率密度(存在確率)が在るのは $x > 0$(体重は正の値をとる。)という当然の条件に 矛盾するという事である。

厳密にいえば、データの存在範囲が$-\infty$ から $+\infty$ までの全範囲、即ち定義域が、$-\infty < x < +\infty$ で無ければ正規分布に従うことは無い。よって、正の値しか取り得ない何らかの「大きさ」の分布は、正確には正規分布に従うことは無い。然るに現在、そのような分布を正規分布に従うとみなして近似している統計が多い。それは正規分布近似の誤適用であると言えよう。

## ４． 対数正規分布

人(生物)の身長や体重など何らかの量の「大きさ」は正の値しかとらないので、その分布は厳密には正規分布になり得ないのであれば、どのような分布に従うのであろうか。例えば図5-1のようになるであろう。つまり、平均値を中心に左右対称なグラフにはならない。$x$ の値は小さくても、限界値が0であるのに対し、$x$ の値が大きくなる分には、理論上、限界値は存在しないからである。

図5-1のグラフは、対数正規分布のグラフである。横軸を対数目盛にすれば、正規分布のグラフと同じ形になる。すなわち、この分布に従う確率変数 $x$ の対数の値を横軸にとったとき、正規分布になる。

平均値$\mu$, 標準偏差$\sigma$の対数正規分布は、記号 $\Lambda(\mu, \sigma^2)$ であらわす。

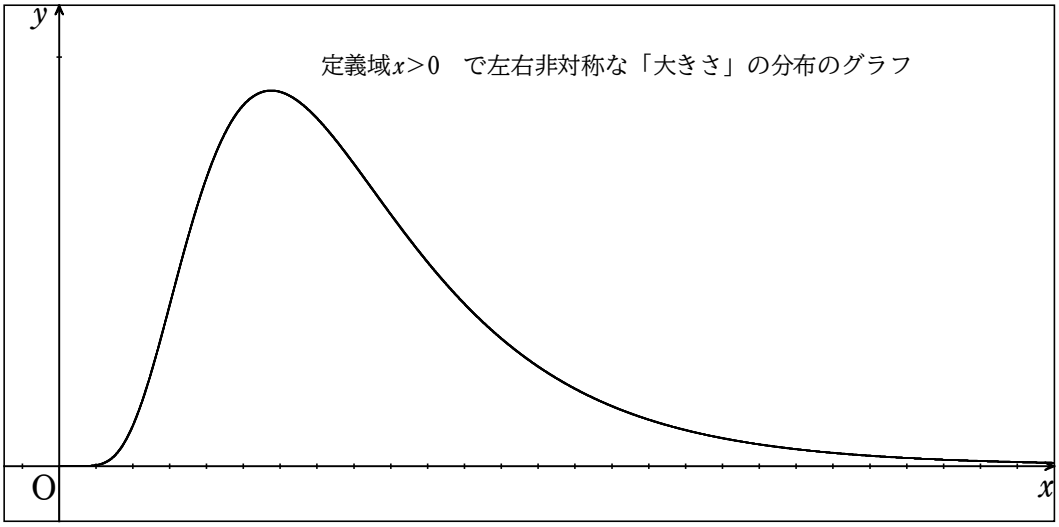


**図５－1　対数正規分布　$\Lambda(\mu, \sigma^2)$の概形**

対数正規分布に従うと考えられている分布には、鉱物塊(岩、石、砂、土)の大きさ(寸法、質量)、油田の規模、材料の寿命、(資本主義社会での)国民の所得の分布、法人の規模の分布　などがある。これらの分布は図5-1のような最頻値(ピーク)に対して左右非対称な分布になる。格差社会の現在は、国民の所得の分布、法人の規模の分布などは、もっと極端に左右非対称で $x$ の正の部分の裾野の長い図5-2　のようになるであろう。これは、平均値$\mu$に対して、標準偏差$\sigma$が大きな対数正規分布である。

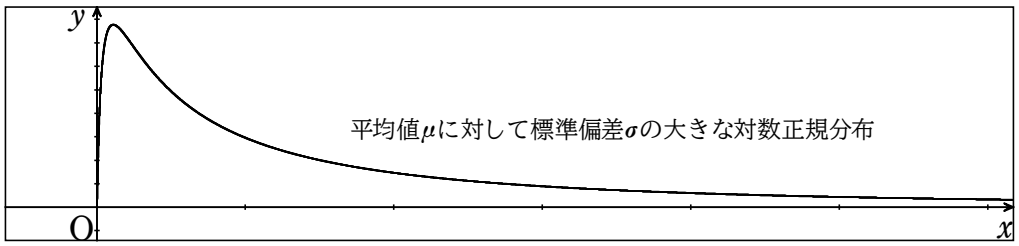


**図５－2　対数正規分布 $\Lambda(\mu, \sigma^2)$ …平均値$\mu$に比べて$\sigma$が大きい場合**

図6は、もとの関数の平均値$\mu$を同じ値1としたときに、もとの関数の標準偏差$\sigma$を、$\sigma=0.4$ ，0.2 ，0.1 と変えた場合の対数正規分布のグラフの形である。標準偏差$\sigma$の値が小さくなるに従い、分布の幅が狭くなって行くのは当然であるが、その形は、左右対称な正規分布に近づいていく。すなわち、対数正規分布のグラフの形は、平均値$\mu$に比べて標準偏差$\sigma$の値が小さければ小さいほど、幅は狭くなり、平均値を中心に左右対称な正規分布の形に近づいて、正規分布で近似できるようになっていく。グラフから $\sigma=0.1$ くらいに小さくなると、かなり正規分布に近くなる。図6から、標準偏差$\sigma$の値が平均値$\mu$の $\frac{1}{10}$ 以下くらいから、対数正規分布は、正規分布で近似できるようになると考えられる。

何故なら $\sigma=0.1\mu$ ならば、$\mu=10\sigma$である。よって $N(\mu, \sigma^2)= N(\mu, (0.1\mu)^2)$に於いて 、

$P(x\leqq 0) = P(x\leqq \mu-\mu)=P(x\leqq \mu-10\sigma) \fallingdotseq 0$

であるから正規分布に近似した場合、この分布は負の値を取る可能性は極めて低く（ほぼ0）、最頻値を軸に左右対称なグラフと見なせるからである。

※＜注＞　対数正規分布を正規分布で近似する場合の平均値は、元の平均値とは異なる。

図6のグラフの平均値は1になっていない。

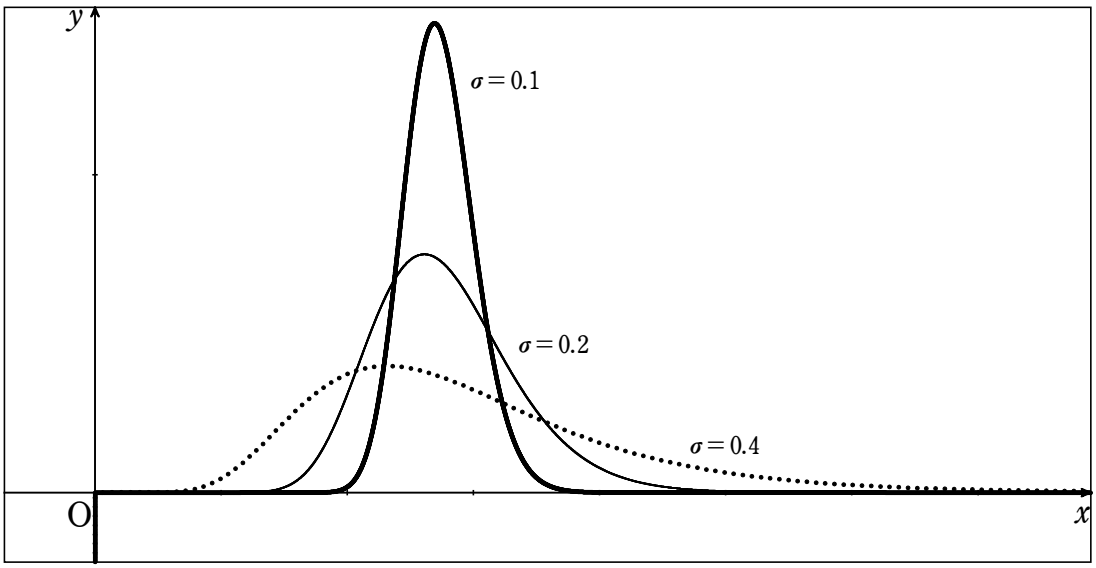


**図６ 平均値$\mu$＝１としたときの、標準偏差$\sigma$の値による対数正規分布の概形**

従来、正規分布に従うと考えられてきた、正の値しか取り得ない分布の多く、例えば生物の体の大きさ、（身長、体重、体の各部位の大きさ等）は、実は対数正規分布に従うのであるが、正規分布で近似していたに過ぎない

と考えられる。

完全に正規分布に従う分布は、測定における偶然誤差や理想気体の（方向の正負を考えた）速度分布くらいしか考え付かない。実際に、ほぼ完全に正規分布に従う分布はあまり無いのではないかと考えられる。

## ５． 正規分布に身長は従うが、体重は従わないように見える理由

先に論じたように、体重が正規分布に従うとすると矛盾が生じる。正規分布の例として、よく同年代、同性の身長が挙げられるが、体重の例はあまり挙げられない。それは体重がきれいに正規分布に従わないからであろう。その理由は、身長と体重の平均値 $\mu$ と標準偏差 $\sigma$ を比べてみればわかる。

政府統計ポータルサイト[e-Stat]によると、2018年の男子の身長と体重の平均値 $\mu$ と標準偏差$\sigma$は、表1-1の通りである。はじめに、年齢は限定せずに全年代の値とした。 これを対数正規分布近似でグラフにすると図7-1のようになる。

**表1-1.日本男性の身長と体重の平均と標準偏差**

| | 平均値$\mu$ | 標準偏差$\sigma$ | $\frac{\sigma}{\mu}$ |
|---|---|---|---|
| 身長$[cm]$ | 162.4 | 18.6 | 0.115 |
| 体重$[kg]$ | 61.8 | 17.6 | 0.285 |

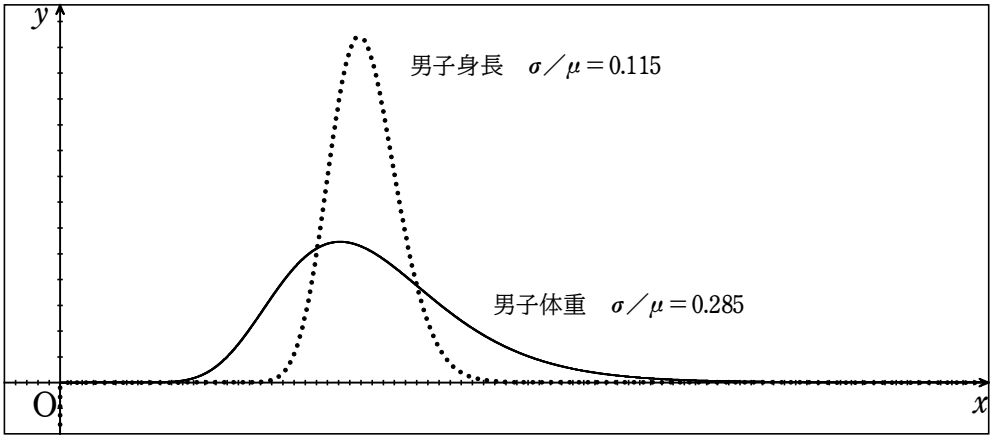


**図７－１　もとの平均値$\mu$＝1とした場合の男子の身長と体重の $\frac{\sigma}{\mu}$ の値の違いによる対数正規分布近似のグラフ**

図7-1は、身長と体重のグラフを比較するために、共に平均値$\mu$＝1として、標準偏差$\sigma$の値を換算して描いた対数正規分布の近似曲線である。グラフからも明らかなように、体重のほうの近似曲線は、明らかに左右非対称である。$\frac{\sigma}{\mu}$ の値がより小さい身長のほうが、体重よりも正規分布に近い形をしている。ただ、これは、年齢を限定しない全年齢でのデータなので、まだ、完全に左右対称の曲線からはやや遠く、やっと正規分布近似できるか出来ないかの程度であると考えられる。

更に年齢も17歳とかに限定すれば標準偏差はより小さくなるので、正規分布により近似し易くなる。

年齢を限定せずに全年代とした上記の場合と、一定の年齢（17歳としてみる。）に限定した場合の男性の身長

の対数正規分布近似のグラフを比べてみる。

**表1-2.日本男性の身長；全体と17歳の平均と標準偏差**

| | 平均値$\mu$ | 標準偏差$\sigma$ | $\frac{\sigma}{\mu}$ |
|---|---|---|---|
| 全体身長$[cm]$ | 162.4 | 18.6 | 0.115 |
| 17歳身長$[cm]$ | 169.2 | 5.8 | 0.034 |

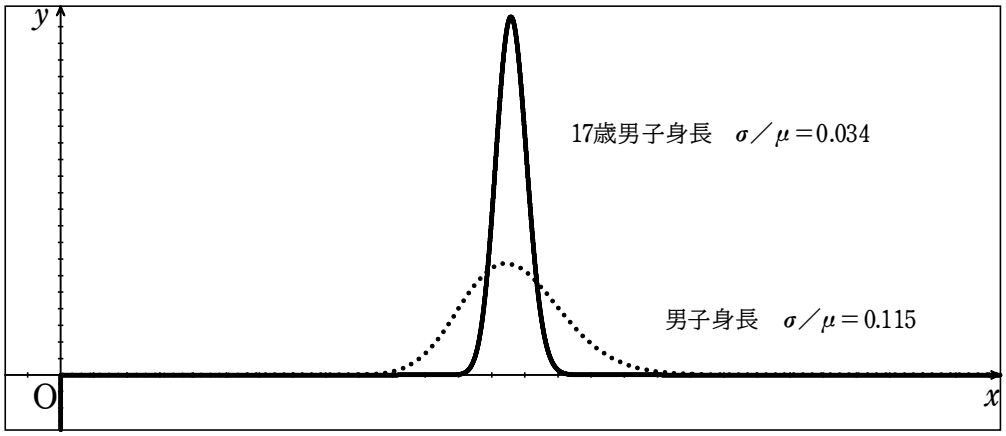


**図７－２　もとの平均値$\mu$を1とした場合の男子の身長と17歳男子の身長の$\frac{\sigma}{\mu}$の値の違いによる対数正規分布近似のグラフ**

※図7-2は、図7-1よりも横軸のスケールを伸ばしてある。

17歳（＝一定年齢）の男子の身長はかなりいい近似で正規分布に近似できる。標準偏差$\sigma$は平均値$\mu$の0.034倍、$\frac{1}{30}$程度であるので、十分に正規分布に近似できると考えられる。

以上、結論として、身長も体重も、当然のことながら正の値しかとらないので、正規分布よりも対数正規分布で近似するほうが適切であると考えられる。身長の標準偏差 $\sigma$ は平均値 $\mu$ の10分の1レベル（以下）であるので、正規分布に近似できるが、体重の標準偏差 $\sigma$ は、平均値 $\mu$ の10分の1より大きく、正規分布とは見なしにくく、正規分布で近似するのには若干の無理がある。

## ６．　正規分布と対数正規分布の対比

平均値 $\mu$ と標準偏差 $\sigma$ が同じ　対数正規分布 :$\Lambda(\mu, \delta^2)$　と　正規分布: $N(\mu, \delta^2)$　のグラフを比べてみる。図8-1, 図8-2　は、$x$軸 ，　$y$軸　それぞれ同じスケールにしてある。

対数正規分布の底は一般には自然対数の底 $e$（＝2.718281828……）を用いるが、以後の議論ではわかり易くする為に、便宜上、底は2を用いる事とする。底が自然対数の底　$e$　のときに $\Lambda(\mu, \delta^2)$　の記号で表すので、底を2にした場合、$\Lambda(\mu, \sigma^2)$と表記することとする。

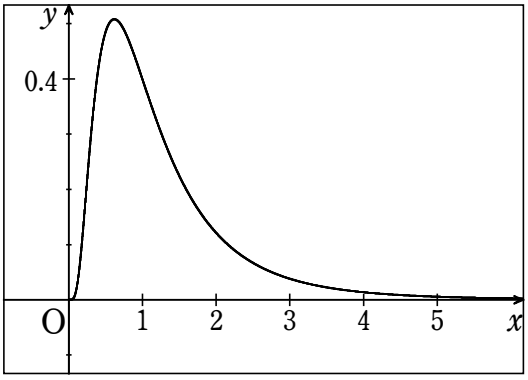


**図８－１対数正規分布　$\Lambda'(0,1^2)$**
**横軸 $x$ は通常の等差目盛：$x=2^X$**

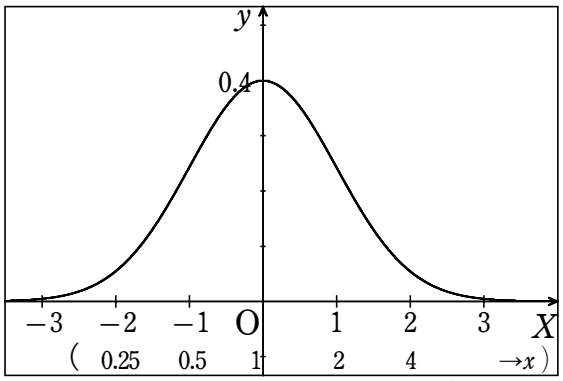


**図８－２対数正規分布 $\Lambda'(0,1^2)$**
**横軸 $X$ は対数表示：$X=\log_2 x$**

図8-1と図8-2のグラフは共に同じ 平均$\mu=0$、標準偏差$\sigma=1$の対数正規分布 $\Lambda'(0,1^2)$である。横軸、縦軸共に同じスケールで描いている。ただし図8-2の場合、横軸は対数目盛になっているので、横軸を$X$ とした。同じグラフなのに形が違うのは、図8-2は横軸が対数目盛 $X$だからである。対数正規分布 :$\Lambda'(\mu,\delta^2)$ のグラフの横軸 $x$を対数目盛 $X$にすればグラフは正規分布の形になる。

図8-1の横軸 $x$ と 図8-2の横軸 $X$ は、 $X=\log_2 x$, つまり、 $x=2^X$ の関係が成り立つ。表2.の通りである。 図8-2の横軸$X$を図8-1の横軸 $x$と同じ基準で目盛れば、図8-2のグラフの $x$ 軸の目盛の下の段に( $\rightarrow x$ )で目盛ったようになる。

**表2．正規分布の横軸 $x$ と対数正規分布の横軸 $X$ の対応表**

| $x$ | 0.25 | 0.5 | 1 | 2 | 4 |
|---|---|---|---|---|---|
| $X=\log_2 x$ | −2 | −1 | 0 | 1 | 2 |

| $X$ | −3 | −2 | −1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| $x=2^X$ | 0.125 | 0.25 | 0.5 | 1 | 2 | 4 | 8 |

対数正規分布の優れた点の一つは、定義域が正の値の分布を $-\infty$ から $+\infty$ の分布に変換することである。即ち $0<x$ の定義域を $-\infty<X<+\infty$ に変換するのである。この変換によって、$0<x$ が定義域の対数正規分布のデータを正規分布に変換できる。また、確率変数の標本の積の分布は対数をとると和の形に変形できるので、十分に大きい確率変数の積は、中心極限定理から、対数正規分布に近づくと考えられる。

※<注>　0以下の値の対数値は無いので、当然、対数正規分布は負の値を取るデータには適応できない。

以下、議論を簡単にするために、正規分布 $N(2,1^2)$ と対数正規分布 $\Lambda'(1,1^2)$ で話を進める。

先ずは図9-1の 正規分布 $N(2,1^2)$ で区間 $1\leqq x\leqq 3$ を考える。この区間の両端の境界値の平均値は、相加平均 $\frac{1+3}{2}=2$ となり、この正規分布全体の平均値(かつ最頻値かつ中央値)と一致する。

次に図9-2の対数正規分布 $\Lambda'(1,1^2)$ で区間 $0\leqq X\leqq 2$を考える。図9-2は、横軸 $X=\log_2 x$ の目盛のグラフである。対応表2より区間 $0\leqq X\leqq 2$ は、区間 $1\leqq x\leqq 4$ である。この対数正規分布の平均値 $x=2$ ( $X=1$ )は、この区間の両端の境界値 $x=$ 1, 4 ( $X=$ 0 , 2 ) の相乗平均 $x=\sqrt{1\cdot 4}=2$ ($X=1$ )となっている。

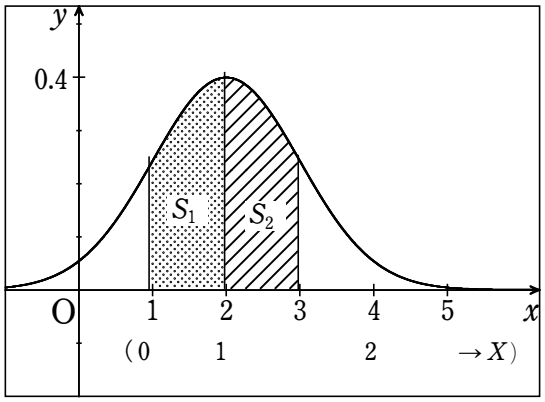


**図９－１正規分布 $N(2,1^2)$**

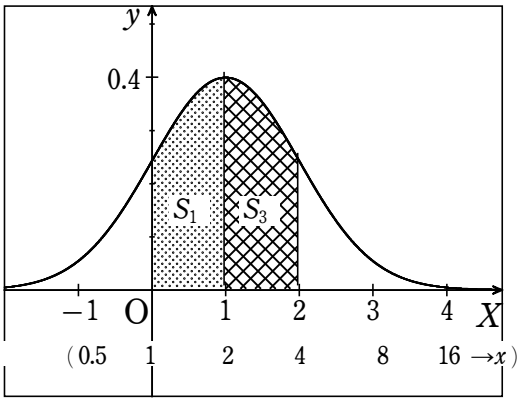


**図９－２対数正規分布 $\Lambda'(1,1^2)$**

一般に相乗平均値は相加平均値以下となるが、話を簡単にするために、平均値は同じ値として議論を進める。図9-1の正規分布で $S_1=S_2$ は統計上の常識である。即ち正規分布の平均値を挟んで両側に同じ幅（この場合、標準偏差 $\sigma=1$ ）だけ取った領域の確率は同じである。式で表すと次のようになる。

$P(1 \leqq x \leqq 2) = P(2 \leqq x \leqq 3)$ （ $\fallingdotseq 0.3413$ ）…①

図9-2の対数正規分布で $S_1=S_3$ という事は、

$P(0 \leqq X \leqq 1) = P(1 \leqq X \leqq 2)$（ $\fallingdotseq 0.3413$ ）

という事である。$x=2^X$ より、$x$ の式に書き換えると

$P(1 \leqq x \leqq 2) = P(2 \leqq x \leqq 4)$ （ $\fallingdotseq 0.3413$ ）…②

という事になる。これは、横軸を通常の目盛に直すと、平均値を挟んで両側の確率が同じ領域の幅は、領域の右側が領域の左側よりも長くなる（この例の場合2倍になっている。）という事である。

図11-1が相当する。

①と②の左辺は共に $P(1 \leqq x \leqq 2)$ で $S_1$ となり同じであるが、右辺は

正規分布①では $P(2 \leqq x \leqq 3) = S_2$ であり、

対数正規分布②では $P(2 \leqq x \leqq 4) = P(1 \leqq X \leqq 2) = S_3$

であり異なる。通常の正規分布の①に関して説明は不要であるので、対数正規分布の②を解釈する。

例として、一定の領域に棲むある種の生物の体長による個体数を考える。生物の種は何でもいい。その生物の年齢（生まれた時からの歳月）は限定しないものとすると、（小さなものから大きなものまでで分散が増えるから）標準偏差 $\sigma$ は大きくなる。（但し、昆虫のように変態する生物は大きさの定義、比較が難しいので除く。）先の議論から、この分布は正規分布とはかなりかけ離れた図5-1や図5-2のようなグラフになるであろう。この分布が対数正規分布に従うとし、横軸を対数表示 にすると左右対称な図9-2のようなグラフになる。このように、対数正規分布では、横軸を通常の目盛り（等差数列）とした場合、左右対称とはならず、平均値と最頻値[mode:モード、ピーク]も重ならない図10のようなグラフになる。

平均値は最頻値[mode:モード、ピーク]よりも大きな値（最頻値よりも右側）になり、その結果、平均値 $\mu$ からの偏差が同じ値: $a\sigma$ だけ離れても平均値 $\mu$ と平均値から $a\sigma$ だけ離れた値とに挟まれる部分の面積（＝その領域の存在確率）は図10のように、平均値よりも左側 の$S_l$ のほうが、平均よりも右側 の$S_r$ よりも大きくなる。

これは生物の体長や体重のように、正の値しか取らないものの分布の特徴としては、ある意味当然の結果と考

えられる。つまり、大きい個体数は少なく、小さい個体数は多いという自然界ではごく自然の事である。

まとめると、正の値しか取らない「大きさ」の分布に関しては、これまでの通常の解釈①、図9-1のように、相加平均の左右での幅が同じ区間での存在確率が等しいという解釈は誤りであり、解釈②、相乗平均の左右での対数の幅が同じ区間での存在確率が等しいと言う解釈のほうが、現実をより正確に近似できると考え得る。

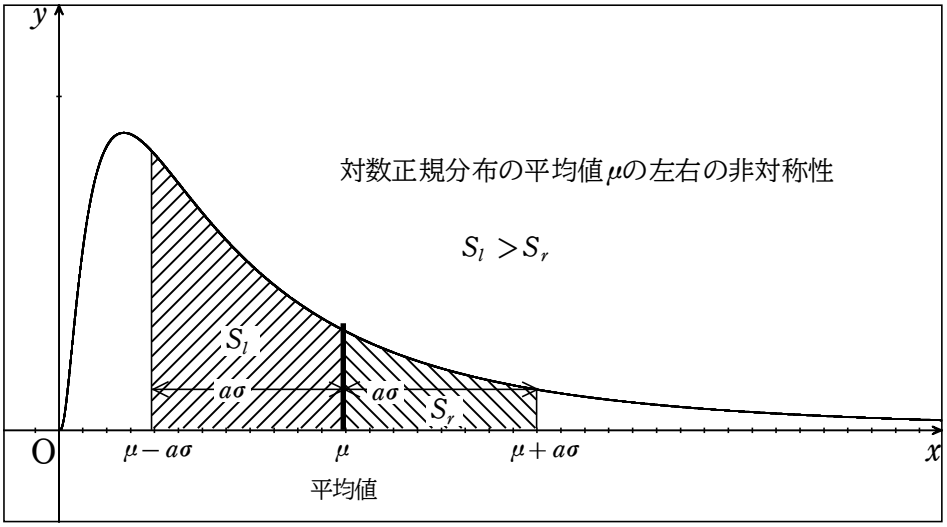


**図10. 対数正規分布 $\Lambda'(\mu, \sigma^2)$の平均値$\mu$±$\sigma$区間の非対称性**

例えばフナなりサンマなりの魚を考えてみる。成長してある程度の大きさの安定した「成魚」（「成人」、「成虫」のように大人の魚を指す）の個体数はある程度多いが、平均値 $\mu$ より更に大きい図10のグラフの平均値 $\mu$ より右側の個体数は、右側に行くにつれてどんどん減っていく。一方、成魚よりも小さな個体数は非常に多く、平均値 $\mu$ からの偏差が同じ $a\sigma$ でも、図10のように、平均値 $\mu$ よりも左側の確率密度関数の値は大きくなる。だから図10のような非対称なグラフになる。

それ故、図9-2で、$S_1 = S_3$ すなわち、横軸が対数目盛で描いた時の平均値の左右の領域の面積が同じ確率と考えることが可能である。逆に、一般に受け入れられて使われている図9-1は、実際の個体数をうまく近似できないと考える。

これを平均体長が2mの魚の例で考える。同じ偏差$\sigma$＝1mでも、その平均値2mの左側では、平均値2mの右側よりも確率密度（存在確率）が大きいので、体長1m～2mの魚の個体数は体長2m～3mの個体数よりも多く（これが図10の示すところである。）、体長2m～4mの魚の個体数と同じであるという結論である。即ち、図9-2の示すように、横軸が対数目盛のグラフでの平均値の左右での同じ幅 $a\sigma$ 内の個体数を同じと考える理論である。通常の相加平均に慣れている人たちには感覚的に馴染まない理論であろうが、 2.相加平均、相乗平均　のところで説明したように、等比数列的（指数関数的、倍数的）自然界では、対数正規分布での確率計算のほうが理に適うと考える。

平均値の左側と、平均の右側の面積（確率）が等しくなるのは、図11－1のような場合である。すなわち図11－1で、標準偏差$\sigma$の係数 $a, b$ について $b < a$ になる。即ち平均値$\mu$の左側の区間よりも、右側の区間のほうが長くなって $S_L = S_R$ となる。図11-1のグラフからも明らかであろう。このグラフの横軸を対数に直せば、図11-2 のような正規分布の形のグラフとなる。図11-1と図11-2は同じ対数正規分布のグラフである。

$S_L = S_R$ で 、$S_L$, $S_R$ はそれぞれ同じ確率（面積）である。

図11-2 は、図9-2と同じグラフである。図1－1、図1－2の相乗平均の例で説明した、両端の相乗平均$\mu$を挟んだ左右で存在確率が同じという事である。 $x$ 軸 方向の幅は平均値の左側（値が小さい領域）よりも右側（値の大きい領域）のほうが長くなる。

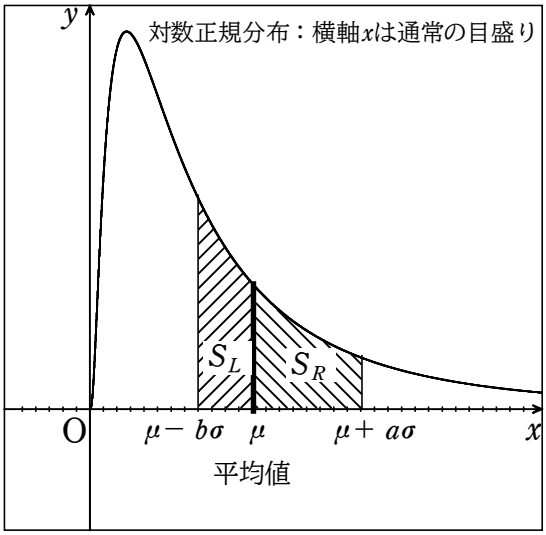


**図11－1　対数正規分布 $\Lambda'(\mu, \sigma^2)$**
**平均値の左右で同じ確率領域**
**横軸 $x$ は通常の等差目盛**

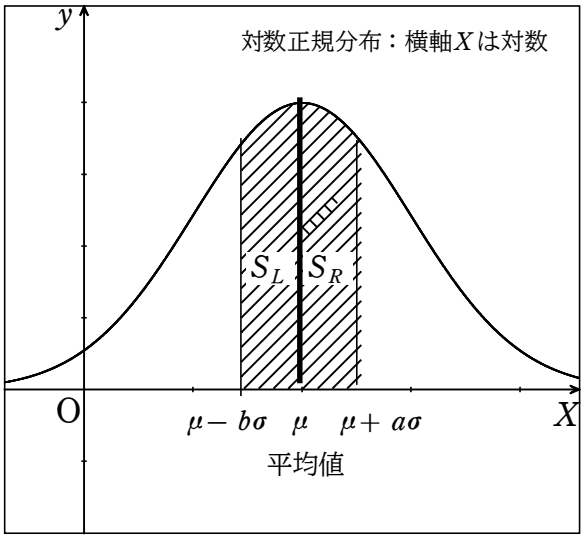


**図11－2　対数正規分布 $\Lambda'(\mu, \sigma^2)$**
**平均値の左右で同じ確率領域**
**横軸 $X=\log_2 x$　（対数目盛）**

## ７．　正規分布と対数正規分布の使い分け

正規分布と対数正規分布の関係は、相加平均と相乗平均の関係と同様に、自然現象を等差数列的・線形に捉えるのか（和の世界）、等比数列的・指数関数的に捉えるのか（積の世界）の違いである。

偶然誤差に関するガウスの3つの公理のうち最初の2つの公理の「誤差」を「偏差」に言い換えてみる。すなわち、

1.大きさの等しい正と負の偏差は等しい確率で生じる。
2.小さい偏差は大きい偏差より起こりやすい。

とする。これを仮に「**偏差の公理**」と呼ぶこととする。

正規分布はこの2つの「偏差の公理」が成り立つ。一方、対数正規分布は、正規分布のデータ $x$ に対して、$X=\log_a x$　（ただし底 $a$ は、$1<a$　を満たす適当な値とする。）と対数に変換したときに $X$ は「偏差の公理」が成り立つ。

正規分布と対数正規分布のどちらかの分布を選んで適用するとするならば、その使い分けは単純である。基本的に、正負の値を取り得るデータの分布には正規分布、正の値しか取らない「大きさ」等のデータの分布には対数正規分布を用いれば良いと考え得る。そもそも0以下の値の対数は存在しないので、対数正規分布は0以下の値を取り得るデータの確率モデルには使えない。勿論、どちらにも従わないその他の確率分布のモデルを用いたほうが良い統計データもある。

I. 正規分布近似を適用する「正負の値を取り得るデータの分布」は、
基本、平均値が0で、最頻値[mode、ピーク]と中央値[median]も0で、その平均値0を中心に線対称の形のグラフになる。例として
・測定器具による偶然の測定誤差
・特定絶対温度での気体分子の熱運動の一定方向の速度成分(正負の値を考慮)
などが考えられる。
II. 対数正規分布近似を適用する「正の値しか取らないデータの分布」は、
・生物の体長、体重、各部位の長さ・質量など「大きさの分布」が考えられる。
・無生物でも、例えば川の長さ、山の高さ、なども対数正規分布に近似出来るであろうと考える。
個人的には、対数正規分布こそ、正の値しか取り得ない倍数的(指数関数的、等比数列的)世界とみなせる自然界の「大きさ」を最も的確に記述する確率モデルであると考える。
先に述べたように、偶然の測定誤差は、正規分布(=ガウス分布)の理論の元になった統計量であるので、正規分布に従う代表例である。しかしこれは、等差数列的に目盛られた測定機器での偶然誤差の測定だからであろうと考える。例えば酸性度pHの測定でも、測定の偶然誤差は、正規分布に従うであろうが、もともとpHは水素イオン濃度の対数の値(のマイナスの符号を取り去った値)なので、pHの測定誤差が正規分布に従うという事は、水素イオン濃度の測定誤差は対数正規分布に従うということになる。
また、(理想)気体分子の熱運動の方向を考えた速度の分布も正規分布に従うが、方向を考えない速さ(速度の絶対値)の分布は対数正規分布に従うと考えられる。
正の値しか取らない分布であっても、元の分布の標準偏差$\sigma$の値が平均値$\mu$の値よりも十分に小さい場合($\sigma \ll \mu$)、この分布は正規分布に近似できる。正の値しか取り得ないデータの分布で、従来、正規分布に従うものとして扱われてきた統計量は非常に多い。代表例に、先に議論した同性・同年齢の人の身長がある。しかしこれらも、対数正規分布で近似したほうが、より適切な近似になるであろうと考える。一般に正の値しかとらない「大きさ」の分布は、従来のように正規分布で近似するよりも、対数正規分布で近似したほうがより良い近似が得られるであろうと考える。現在、何でもかんでも正規分布に近似して分析する傾向があるが、正の値しか取り得ない統計量の分布は、対数正規分布が基本である。正規分布と決め付けずに、対数正規分布近似も試みて比較して、現実をより正確に記述しているほうを選ぶべきである。(対数正規分布のほうを選ぶことになるであろう。)

## 8. 今後の課題

以上で議論したように、何らかの大きさのように、正の値しかとり得ないデータの分布は、正規分布よりも対数正規分布のほうがより適切に近似できると考える。本稿では、正の値しかとらないデータの対数正規分布での近似の有効性について論じたが、実際のデータでモデリングしての議論が足りなかった。実際に大量のデータを用いて対数正規分布での近似を行い、その有効性の検討を行いたい。その場合、大量のデータの度数分布表の作成が必要となる。1種類のデータだけでも、手間暇がかかるので、項目数を絞って、自らデータを取るか、またはどこかの利用可能なデータを用いて検証してみたいと考える。特に、ある領域の(例えば日本の)川の長さ、山の高さ、ある地点での風速(の絶対値)、などの分布も対数正規分布に近似できる可能性が高いと考えているので、検証したい。

## 主要参考文献

[1]竹内啓（1985）数理統計学　データ解析の方法　東洋経済

[2]守谷栄一（1974）詳解演習　数理統計学　日本理工出版会

[3]原島鮮（1966）熱力学・統計力学　培風館

[4]ティ.ア.アゲキャン　本田勝．幾志新吉　訳（1969）誤差論の基礎　総合科学出版

## 主要参考Webサイト

[1]ガウス分布の導出

https://www.eng.niigata-u.ac.jp/~nomoto/7.html

[2]中心極限定理　統計WEB

https://bellcurve.jp/statistics/course/8543.html

[3]石油/天然ガス用語辞典　Weblio

https://www.weblio.jp/content/lognormal+distribution対数正規分布

[4]政府統計ポータルサイト[e-Stat]都道府県別　身長•体重の平均値及び標準偏差

https://www.e-stat.go.jp/dbview?sid=0003146482

[5]Digiland　dB（デジベル）って何？音圧とは？

https://info.shimamura.co.jp/digital/knowledge/2014/03/21161

## 謝辞

本稿は、会津大学短期大学部 若林達司先生のお勧めで書かせて戴きました。心から感謝申し上げます。この論文の作成によって、学生時代から疑問に思っていた正規分布の欠点、矛盾点を指摘し、正の値しかとらない「大きさ」の分布は、正規分布に換えて、対数正規分布を適用すべきであるという持論に関して研究し、洞察を深め、ある程度までまとめることが出来ました。