七つの正規分布－正規分布はどこから来たのか？－

○椎名乾平　（早稲田大学教育学部）

Key words: 正規分布の起源，統計学史，心理学的測定法

心理学の様々な分野で正規分布がデフォルトの分布として仮定され使用されている．根拠がある場合もあろうが，特に根拠もなく過去の習慣や**願望**によって使用されている場合も多い．本稿はこのような状況に対し，ある種の示唆を与えるのを目的とする．その論点は以下の２点にまとめられる．

論点１　歴史的に見た場合，正規分布は約３００年かけて，各時代の天才級の人たちによって再発見・再解釈され続けてきた（表１）．数式が同じになる以上，再発見再解釈には二次的意味しかないと考えるのは大きな誤りである．なぜなら，「再発見」は，異なる文脈で切実な**「具体的問題」**を解くために，異なる**「根拠・前提」**を用いて行われて来たからである．本論文では７つの異なる**「具体的問題・根拠・前提」**を示し（最後の一つは「観察」であるが），その「根拠・前提」から正規分布がどのように「導出」されたのかを概観する．

論点２　過去の様々な正規分布導出をその**「問題」「根拠・前提」**と共に展望すると，心理学で出現する様々な正規分布にどのような根拠や裏付けがあるのかという問題に，洞察と指針を与えてくれる可能性がある．すなわち，正規分布はそれがどのように導出されたかという観点から見ると決して一枚板ではなく，様々な変化形がある．そこで，心理学で現れてくる正規分布がどの変化形に近いのかを考察すれば，心理学研究者の持つ暗黙理論・仮定が見えてくるかもしれない．

●7 つの「具体的問題」と正規分布の導出　表１に７つの正規分布についてまとめた．紙面の関係で５）については省略する．**１）賭けの確率**　確率論がギャンブルの研究から始まったのはよく知られているが，この文脈内で 2 項分布が正規分布で近似できることが明らかになった．この正規分布には $\sigma^2 < \mu$ という制約がつくので，一般の正規分布ではない．逆にいえば，分散が平均より大きい場合この正規分布では説明がつかないことになる．**２）天体観測・測量の精度**　天文学・測地学では航海術や政治・行政の要請から正確な測定が要請された．この文脈から最小二乗法や　**測定値＝真値＋誤差**というモデル（後世，古典的テスト理論や分散分析に繋がる）が提案された．Gauss の導出は，データの平均値が最確値であるためには，誤差分布は正規分布でなければならないという（逆転の）論理構成をとっている．元々無関係であった１）と２）の関係に気づき統合したのは Laplace(1810)と言われる．**３）Herschel の法**　１）２）での正規分布の導出は難しいが，この方法は驚くほど簡単で短い．平面上の確率密度 g が原点からの距離の関数であり，また x 軸 y 軸上の密度 f の積で表現できる，すなわち $f(x)f(y) = g(\sqrt{x^2+y^2})$ とする．$y = 0$ とおくと $f(x)f(0) = g(\sqrt{x^2+0^2}) = g(x)$ だから $f(x)f(0) = g(x)$ でなければならない．これより $g(\sqrt{x^2+y^2}) = f(\sqrt{x^2+y^2})f(0)$ となる．元の式に代入すると $f(x)f(y) = f(\sqrt{x^2+y^2})f(0)$ となり，よって $\ln f(x)/f(0) + \ln f(y)/f(0) = \ln f(\sqrt{x^2+y^2})/f(0)$ この関係を満たす $f(x)$ は $\ln f(x)/f(0) = ax^2$ しかないので，$f(x) = Ce^{ax^2}$ が導き出される．後は定数を決定すればよい．

４）確率過程　心理学では時間と共に変化する正規分布を使用する例は少ない．しかし，時系列分析や力学系のモデルの発展とともに今後注目されるものと予想される．理論的には１）の後裔とも考えられる．**６）中心極限定理**も１）の後裔であるが，独立性の仮定，同分布の仮定を緩める研究が進んでおり，正規分布が出現可の文脈は理論的には拡大している．

７）経験的観察　現在でもデータが正規分布する理論的保証を与えるのは難しい．実データをとり確証する必要がある．一方，根源誤差（Hagen,1837）が積み重なって，測定値が正規分布するという**思想**には様々なバージョンがある．これは１）の後裔であると言えよう．しかし３）５）の正規分布は根源誤差の仮定と無縁と思われる．

平均値の解釈の問題　物理的測定値の場合に，平均値が真値の最確値（推定値）となるのは２）で保証される．一方，Quételet（1846）は社会的測定値と物理的測定値のアナロジー（Stigler, p.221）を立て，社会的・生物学的カテゴリー内のちらばりと（例えば，スコットランド兵の胸囲），単一の物理的対象の多数回測定から得られるデータのちらばりとを（根源誤差を仮定すれば同じ正規分布するのだから）同等のものと主張する．この思想から真値に相当する「平均人」という概念が生まれたのはよく知られているが，当時から相当に批判されて来たにも関わらず現在でも仮説構成体や潜在変数という形で健在と思える．平均身長あるいは平均（尺度）得点のようなものが「実体」なのかあるいは仮説構成体なのかは，データの正規性からだけでは決定できないと考えるべきである．また，潜在的な仮説構成体に対して正規分布を仮定するのは常に危険な行為と考えるべきである．

表１　七つの正規分布

具体的問題	時期	主要人物	根拠・前提・解法のコメント	特徴
１）賭けの確率	1730-1800	De Moivre, Laplace	2 項分布の正規近似	μ，σ^2 は自由に動けない
２）天体観測・測量の精度	1800 前後	Gauss, Legendre	最小二乗法の基礎付け	証明が循環論的
３）平面・空間上の散らばり	1850 前後	Herschel, Ellis, Maxwell	関数方程式	多次元分布から 1 次元正規分布を導出
４）ブラウン運動と株価変動	1900 前後	Einstein, Bachelier	確率過程（時間軸の導入）	時間と共に発展する正規分布
５）情報理論との連携	1950 前後	Jaynes	Maximum entropy	ベイジアン
６）確率変数の和の分布	1900-1950	Lyapunov, Lévy, Lindeberg, Feller	中心極限定理の数学的発展	独立同分布でなくとも正規分布がでることがある
７）経験的観察	1818-	Bessell, Quételet, Galton	物理測定　生物測定　社会測定　遺伝	様々な現象が正規分布に従うことを発見