

Bernstein-von Mises 定理の歴史について*

吉村 有博

要旨

本論文は Bernstein-von Mises 定理に関する古典的な諸結果を概観し、現代的視点からその歴史的意義を明らかにするものである。具体的には定理に関連する書誌情報、定理の形式、当時の動機、応用例を整理し、頻度論統計学との関連を概観する。本論文の主要な貢献は定理の古典的諸結果の歴史的意義と、拡張された現代の定理との関連の解明である。特に、先駆的貢献者の間で事後分布の漸近正規性を導く動機は大きく異なり、その一部は信頼区間の正当化が動機に関連したことを指摘する。さらに、定理の発展には中心極限定理の発展や確率の公理化の運動が強く影響したことを指摘する。加えて、尤度関数の任意性、事前分布の特定化からの独立性、そして区間を代用する可能性の起源はベイズ統計学史の極めて初期に遡れることを指摘する。

キーワード：Bernstein-von Mises 定理、統計学史、確率論史、ベイズ推論、漸近正規性

1. 序論

近年の統計学や計量経済学において Bernstein-von Mises 定理（以後、BvM 定理）の研究が注目されている。この定理は、観測値の数あるいは情報量が多くなると、未知のパラメータのベイズ事後分布が正規分布に近づくことを示した結果あるいは現象の総称である。典型的に知られる定理の特徴は以下に集約できる：標本サイズが大きいとき、有限次元パラメータの事後分布は多変量正規分布に従い、その平均は最尤推定値、分散共分散行列は Fisher 情報行列の逆行列となり、事前分布に依存しない (van der Vaart (1998) など)。まさにこれは中心極限定理のベイジアンアナロジーに相当し、ベイズ推論における基礎的な役割を果たす。特に、正規性の帰結は事後分布の計算を容易にし、また効率性の議論を明確化する。正規性に対する尤度関数の任意性は応用範囲の拡大に貢献する。また、事前分布からの独立性は事前分布の選択の論争を緩和する根拠を与える。とりわけベイズ

* 論文を改善する上で有益なコメントを頂いた小田秀典教授、ならびに令和 2 年度京都産業大学経済学部研究会の参加者の方々に感謝を申し上げます。内容および翻訳における誤りの責任は全て著者に帰する。

アン-頻度論間の分散の一致は興味深い点で、ベイジアン 95%信用区間と頻度論 95%信頼区間が (近似的に) 同一になることを意味する。特に信用区間が仮想的な繰り返し標本に基づく長期的な被覆性質を持ちうる点は重要である。また、ベイジアンと頻度論では哲学や確率解釈が異なるにも関わらず、定理は異なる立場に基づく結果が同一の結論をもたらす可能性を示唆する。さらに定理は信用/信頼区間の計算に互いの計算手法を代用する可能性を示唆し、この帰結は双方の立場にとって有益である (Chernozhukov and Hong (2003) など)。

BvM 定理の原型は 18 世紀後期に P. S. Laplace により発見されたとされ、それ以来確率論や統計学において様々な発展、応用、そして議論を経て、今日に至っている。定理の名称はその発展の貢献者である 20 世紀の数学者 S.N. Bernstein と R. von Mises に由来する。Le Cam (1986) は彼らの貢献を再評価しつつ結果を一般化し、有限次元パラメータをもつ滑らかなモデルに関する BvM 定理の基礎を確立した。これはより理解しやすい形で van der Vaart (1998) などで知られている。20 世紀後半からは無限次元のパラメータを含む問題への定理の拡張が検討されてきた。ここで Freedman (1999) らにより、極めて自然な無限次元モデルの例でネガティブな結果すなわちベイジアンと頻度論の結果が異なる可能性が明らかにされた。この事実は双方の関連の限界を示唆するかのように見えたが、一方で、一部の無限次元モデルや興味あるパラメータが有限次元であるモデルにおいては再びポジティブな結果が引き継がれることが数多く報告されている。20 世紀後半以降の定理の進展は Ghosal and van der Vaart (2017) に整理されている。発見から 2 世紀以上を経てなお定理の拡張研究は進み、頻度論との関連を含めた様々な性質が明らかになりつつある。

BvM 定理の新しい結果に注目が集まる一方で、実は古典的な結果すなわち定理の歴史的な背景については明らかでない部分が存在する。特に、定理名に帰属する Bernstein と von Mises の貢献や背景は、実はよく知られていない。ベイズ推論や逆確率の学史は、主に確率論史の一部として Todhunter (1865)、Stigler (1986)、Hald (2010)、Dale (2012)、Seneta (2013)、Gorroochurn (2016)、Sheynin (2017) に含まれるが、BvM 定理に相当する結果は網羅されておらず、詳しい記述は見られない。特にこれら文献は Laplace には詳しいものの、Bernstein と von Mises に関する記述は驚くほど少なく、彼らの結果やモチベーションは殆ど紹介されていない。とりわけ Le Cam (1986) は現代的枠組みにおいて BvM 定理を再評価した重要文献だが、残念ながら歴史への言及は極めて短く、しかも不正確な書誌情報を含むようである (他にも Le Cam (1953)、Le Cam and Yang (1990))。歴史への言及を含む他の文献として Diaconis and Freedman (1986)、Lehmann and Casella (1998)、Ghosh and Ramamoorthi (2003)、Kleijn and van der Vaart (2012)、Castillo and Nickl (2013) などがあるが、Bernstein と von Mises に関する具体的な結果や動機に関する記述はなく、初出の引用文献すら統一されていない。驚くべきことに、これほど有名な定理にも関わらず、その古典的な諸結果の歴史的背景は整理されていないようである。

ここでの問いは、現代に再び注目される BvM 定理について、その古典的な諸結果がいかなる経緯で発展してきたのかについてである。事後分布の漸近正規性の起源はどこにあり、当時何が問題になり、具体的に何が示されたのであろうか。定理は確率論の数学的拡張においてのみ発展したのか、それとも統計学と関連しながら発展したのか。さらに、古典的に考えられた問題は現代の BvM 定理とどう関連しているのか。既存の歴史文献からはこうした問いへの答えが見えてこないようである。現在において今一度古典的な諸結果の発展過程を概観し、その歴史的意義を考えることには一定の価値があると思われる。

本論文は古典的な BvM 定理の歴史的発展を概観し、現代的視点からその歴史的意義を明らかにするものである。具体的には、1930 年代以前に示された BvM 定理に焦点を当て、関連する書誌情報、当時の動機、定理の形式、応用例、先駆者らの貢献を整理し、頻度論統計学との関連を明らかにする。さらに、現代の定理の進展を踏まえた上で、古典的諸結果の歴史的意義の解明を試みる。特に古典的なモチベーションの解明や頻度論との関わりに対する理解は今後の定理の進展を見る上で重要である。

本論文の主要な貢献は以下の点を明らかにしたことである：(1) 古典的な BvM 定理を導出する動機は貢献者らで大きく異なること、(2) Laplace を始めとする中心極限定理の発展が強く影響したこと、(3) Bernstein と von Mises の定理の背景には確率の公理化の運動があったこと、(4) 既存の学史研究で過小評価された Bayes による近似公式は BvM 定理の萌芽的成果と評価できること。加えて、尤度関数の任意性や事前分布の特定化からの独立性、そして区間の同一性の起源はベイズ統計学史の極めて初期に遡れることを指摘し、それが最初どのように考えられたかを明らかにする。著者の知る限りこれらの点は既存の学史文献で指摘されておらず、BvM 定理に焦点を当て概観することで明確になった知見である。

本論文が扱う BvM 定理は、曖昧さの排除のため、以下を満たす結果に焦点を当てる：ベイズルール（ベイズの定理）を明示し、漸近理論を用いて正確な分散に基づく正規分布を導出し、かつ歴史的に新しいものであること。ただし、事後分布の漸近正規性を導出する背景やモチベーションは時代や貢献者に応じて大きく異なるため、これらの点はこの定義に入れていない。実際、定理名に帰属する Bernstein と von Mises ですら主要な動機は全く異なって見える。本論文は数理的事実に応じてこの定義に合致するものを BvM 定理と見なし、その上で各文献における動機の違いを明らかにする。

古典的な定理の歴史において重要な進展を含む文献は、以下の六つに集約される：萌芽的成果として Bayes (1764, 1765)、Laplace (1774)、明確に新規的な結果として Laplace (1786)、Bienaymé (1838)、von Mises (1919a)、Bernstein (1927) があげられる。ここには主要な動機が必ずしもベイズ推論とは限らないものが含まれることに注意したい。また、Laplace (1812)「確率の解析的理論」は確率論史における重要文献だが、そこで示される BvM 定理には顕著な新規性がないことに加え、既存の学史文献で詳細が知られるため、本論文に関わる結果のみを適宜示す。また、新規的な BvM 定理に該当しないがその成立史に関連する文献として Gauss (1809)、Poisson (1830, 1837)、Pearson and Filon (1898)、

Edgeworth (1908) が挙げられるが、これらについては必要に応じて関連部分に言及するに留める。また、貢献者らの伝記や他分野における貢献については他書で詳細な報告があるため本論文では取り上げない。まとまった伝記は Heyde and Seneta (2001)、Dictionary of Scientific Biography がある。詳しい伝記や参考文献は、例えば Bayes と Laplace は Todhunter (1865)、Stigler (1986)、Hald (2010)、Dale (2012)、Bienaymé は Heyde and Seneta (1977)、Bernstein は Seneta (2013)、Videnskii (2000)、von Mises は Cramér (1953) を参照されたい。ただし、これら文献のほとんどが BvM 定理に関する詳細を含んではいない。

本論文は現代の BvM 定理の理解に動機付けられたものであり、特に“頻度論的な視点から”ベイズ推論の歴史に焦点を当てたものである。特に頻度論的に事後確率の挙動を調べることを一部のベイジアンは魅力的とは思わないかもしれない。この点で本論文の視点は限定的であり、全てのベイジアンから歓迎される歴史的概観ではないであろう。この点は頻度論者にとっても同様で、一部の頻度論者は事後確率を頻度論的に調べることに懐疑的かもしれない。また、本論文は学史研究として不十分なものであり、定理の貢献者らの他の研究との関わりや、確率論や統計学全体の変遷を十分に捉えきれていないことをあらかじめ指摘しておきたい。本論文はあくまで古典的な定理の発展過程に焦点を当て、その現代との繋がりやの解明を目的としたものである。

論文の残りの構成は以下である。第 2 節で貢献者の文献別に BvM 定理の諸結果を概観し、その動機と貢献を明らかにする。第 3 節で現代的な BvM 定理の視点に基づき、古典的結果の歴史的意義をまとめる。第 4 節で結論付ける。

2. 古典的 BvM 定理の歴史

2-1. Bayes (1764, 1765)

BvM 定理の歴史を振り返るとき、既存の学史文献のほとんどにおいて、その最初に P.S. Laplace の名が挙げられる。後に見るように確かに Laplace の成果は疑いようもなく当初の顕著なものであるが、実は T. Bayes もその萌芽的成果を導出していたことはあまり指摘されていない。Bayes の導出した事後確率の近似公式を慎重に検討すると、そこには正規分布に極めて酷似した式が見られる。

ここで取りあげるのは「偶然性の問題に関する解法の論文」と題された Bayes (1764) で、ベイズルールが登場する最初の文献と同一のものである。論文は Bayes によって公刊されたものではなく、1761 年の Bayes の死後、親族の依頼で原稿の整理が託された R. Price によって 1764 年に公刊されたとされる (Todhunter (1865) など)。論文の構成として、前半に Price による前書きがあり、続いて Bayes 本人による本論は“問題”と 2 つの節で構成され、Price による補論で数値例が示される。“問題”では、二項事象の起こる頻度が与えられた下で未知の成功確率が所与の区間に含まれる確率を決定する問題が提示される。

1節では確率の定義と7つの命題と(解析学としては不完全な)証明が示される。2節では有名なビリヤードテーブルに球を落とす問題が提示され、3つの命題と証明が示され、2節最後半において証明なしで3つの“Rule”が示される。ここの命題9が今日知られるベイズルールの特例の場合であり、Rule 2がこれから説明する事後確率の近似公式である。Bayes (1764)ではRule 2は証明なしで与えられたが、その重要性を認識したPriceは翌年Bayes (1765)「偶然性の問題に関する解法の論文におけるRule 2の証明」において、Bayesによる証明を公刊した。

Bayesによる事後確率の近似公式は彼の提示したベイズルールと明確に関連するものである。ここで彼の手によるベイズルールを紹介する。彼は本論冒頭で以下の問題を提示する：“問題：未知の事象が起きた回数と失敗した回数が与えられたとする。このとき、一度の試行でその事象が起こる確率が任意に与えた2つの値の間にある確率を求める”。この問題に対し彼は命題9において、成功が m 回、失敗が $n - m$ 回の計 n 回観測された下で、未知の成功確率 θ が所与の θ_0, θ_1 ($0 \leq \theta_0 < \theta_1 \leq 1$)に含まれる確率を、現代の記法で

$$P(\theta_0 < \theta < \theta_1) = \frac{\int_{\theta_0}^{\theta_1} \frac{n!}{m!(n-m)!} \theta^m (1-\theta)^{n-m} d\theta}{\int_0^1 \frac{n!}{m!(n-m)!} \theta^m (1-\theta)^{n-m} d\theta}$$

であることを示す。これはまさに未知の θ の事前確率を $[0, 1]$ 上の一様分布とした場合のベイズルールと同等のものである。彼は補足(Scholium)において事前確率について議論を行うが、Edwards (1978)によれば、Bayesは事象について前もって何も知らない状況の表現として一様事前分布を仮定するに至ったようである。上式は現代の記法を用いたものであるが、実際にはBayesは積分記号を用いず初等幾何の面積の比として述べている。Bayesは続く命題10で定積分を二項定理から近似し、その計算法をRule 1にまとめる。続くRule 2において具体的な事後確率の近似公式を与える。

BayesによるRule 2のモチベーションは、現代の言葉で言えば、 n が大きい場合の二項分布に基づく尤度関数の定積分の近似である。実際、上式分母の積分はベータ関数の性質より $1/(n+1)$ と簡単になるが、分子は θ の多項式の定積分であり、特に n が大きい時は計算が困難である。ベイズルールは積分比として事後確率の計算ルールを提示するものの、所与の n, m, θ_0, θ_1 に対し計算できないのであれば確率計算として有益ではない。そこで、Bayesは実行可能な公式を解析的近似によって提示するわけである。

命題 (Bayes, 1764, “Rule 2”)

ある事象について何も分かっていないが、 n 回の試行において m 回成功し $n - m$ 回失敗したことが分かっている、 $n - m$ が m より大きいとき、一回の試行において成功する確率 θ が m/n から $m/n + z$ にある確率は

$$\Sigma \leq P\left(\frac{m}{n} < \theta < \frac{m}{n} + z\right) \leq \Sigma \times \frac{1 + 2Ca^m b^{n-m} + 2Ca^m b^{n-m}/n}{1 - 2Ca^m b^{n-m} - 2Ca^m b^{n-m}/n}$$

を満たす。ただし $C = n!/(m!(n-m)!)$, $\tau^2 = n^3/(2m(n-m))$, $a = m/n$, $b = (n-m)/n$,

$$\Sigma = \frac{n+1}{\tau} C \left(\frac{m}{n}\right)^m \left(\frac{n-m}{n}\right)^{n-m} \left(\tau z - \frac{\tau^3 z^3}{3} + \frac{n-2}{2n} \frac{\tau^5 z^5}{5} - \frac{(n-2)(n-4)}{2n \cdot 3n} \frac{\tau^7 z^7}{7} + \dots \right).$$

Rule 2 は Σ に基づく近似の誤差限界を与えるが、このとき $n \rightarrow \infty$ で $P\left(\frac{m}{n} < \theta < \frac{m}{n} + z\right) = \Sigma + o(\Sigma)$ が従うため、 Σ は事後確率の近似の先導項を意味する。ここで、 n が大きければ、 Σ の級数部分は正規分布の核の展開を 0 から z まで定積分したものと漸近同等であることに注意する。実際、簡単な書き換えと Stirling の公式より $n \rightarrow \infty$ において以下である：

$$\Sigma \approx \frac{1}{\sqrt{2\pi m(n-m)/n^3}} \int_0^z e^{-\frac{n^3}{2m(n-m)} t^2} dt.$$

これは、 n が大きければ、実質的に θ の事後分布は $N\left(\frac{m}{n}, \frac{m(n-m)}{n^3}\right)$ で近似されることを意味する。Bayes による Rule 2 の導出は数学的に曖昧で分かりにくい部分を多く含むが、実質的に正規性を導いたことが示唆され、それ自体は画期的な成果と言える。実際、近似に用いられたある対称関数が Bayes (1765, p.298) の図でベルカーブとして描かれるが、これは正規性を強く示唆するものである。ただし、Bayes は厳密な正規分布の核やその定積分 $\sqrt{\pi}$ を示していないため、解析的に見て厳密な漸近正規性と呼ぶことは難しいと思われる。したがって、本論文ではこれを漸近正規性の発見と呼ぶことは控えたい。それにもかかわらず、彼は事後分布の近似的な考察を行うことで、漸近正規に極めて酷似した実行可能な近似公式の導出に成功したと評価できる。

さらに、Bayes (1765) の Article 20 の脚注において、おそらく Price によって、(等裾事後) 信用区間の具体的な構成が見られることは注目すべきである。Rule 2 と近似関数の変曲点に関する考察から、彼らは成功確率 θ が区間 $\frac{m}{n} \pm \frac{1}{\sqrt{2}} \sqrt{\frac{m(n-m)}{n^2(n-1)}}$ に含まれる事後確率は

$1/2$ 、 $\frac{m}{n} \pm \sqrt{\frac{m(n-m)}{n^2(n-1)}}$ に含まれる事後確率は $2/3$ 、 $\frac{m}{n} \pm \sqrt{2} \sqrt{\frac{m(n-m)}{n^2(n-1)}}$ に含まれる事後確率は $5/6$

と示した。ここでは、区間を含む確率に応じて計算された係数 $1/\sqrt{2}, 1, \sqrt{2}$ が分位点の役割で用いられている。定数 $\sqrt{\frac{m(n-m)}{n^2(n-1)}}$ は n と $n-1$ の違いを無視すれば $\sqrt{\frac{m(n-m)}{n^3}}$ で、前述の事後分布の標準偏差に相当する。彼らは数値例として $m = 1,000$ 、 $n - m = 100$ についてそれぞれの区間を具体的に $10/11 \pm 1/163, 10/11 \pm 1/115, 10/11 \pm 1/181$ と示した。

この数値結果と正規近似からの計算を比較したのが表 1 である。まず、所与の水準 α に対し彼らの $\alpha/2$ - 分位点が、標準正規分布のそれを高い精度で近似できている。さらに、確率 $1 - \alpha$ に応じた区間の境界が標準正規分布を基にした区間の境界と極めて高い精度で近似できている (小数点第 2 位まで正確である)。この数値比較は、 n が大きければ Bayes

の近似公式が実質的に正規近似を示していたことを強く示唆する。また Price は信用区間のアイデアを提示しており、それは標準正規分布を用いたものと極めて近いものである。

$1 - \alpha$	Bayes (1765)			正規近似		
	$\alpha/2$ -分位点	下限	上限	$\alpha/2$ -分位点	下限	上限
0.500	0.707	0.9030	0.9152	0.674	0.9032	0.9149
0.667	1.000	0.9004	0.9178	0.967	0.9007	0.9175
0.833	1.414	0.8967	0.9214	1.383	0.8971	0.9211

表1 : Bayes (1765) の“信用区間”と正規近似

驚くべきことに、既存のほとんどの学史文献において、Bayes の Rule 2 は BvM 定理に関連するものとして評価されていないように思われる。とりわけ Hald (2010) は Bayes の Rule 2 と正規分布の関連を明確に指摘したにも関わらず、Bayes の貢献を高く評価しているようには見えない。最近 Sheynin (2017) がおそらく Rule 2 と正規性の関連を指摘しつつ Bayes の貢献を高く評価したものの、信用区間の萌芽を含めた詳細な議論はやはり行っていない。BvM 定理の歴史において Bayes の近似公式に言及すべき根拠は少なくないように思われる：(1) 近似公式の公刊年 1764 年は後述する Laplace (1774, 1786) に先んじること。(2) 事後確率の近似公式に至るモチベーションが Laplace (1786) と大きく変わらないこと。(3) 解析学として不完全ながらも証明がわざわざ(翌年)提出されていること。(4) 近似公式は実質的に正規分布を意味し、その平均と分散も正しいこと。とりわけ、近似に用いられた対称関数はベルカーブとして論文中で図示されるが、その曲線は正規分布と酷似したものである。(5) おそらく Price によって信用区間の概念が数値例と共に明確に導出され、正規近似の場合と極めて近いこと。これはデータを与えれば実際的な応用が可能であり、彼らの方法は逆問題の解法の一つとして完結することを意味する。以上の根拠から、Bayes の近似公式は BvM 定理の歴史において議論なしに省略できるほど無関係のものではないと指摘できる。

Bayes の結果の歴史的重要性は、BvM 定理の萌芽としての事後確率の近似公式である。とりわけ重要なのは、漸近理論を考え、対称関数による近似を見だし、正規分布の核と同等の展開を導き、導出された分散が正確なことである。特に、本来歪んでいるはずの事後分布をモード周りでベルカーブ型の対称関数で近似できることを既に見抜いている点は無視できない事実である。この正確で実行可能な近似により、Price によって信用区間が最初に提示され、それは正規近似による計算と酷似したものである。後に Laplace (1786) によって解析的に事後分布の漸近正規性が示され、Laplace (1812) で Laplace による信用区間が提示されるが、Bayes らのものはそれに少なくとも 20 年先んじる。以上の事実を踏まえれば、Bayes の結果は BvM 定理の歴史において言及する価値は十分にあると思われる。ただし、解析学として厳密な事後分布の漸近正規現象の発見者は、Laplace に譲ら

ざるを得ないようであるが、この点について以降の項で明らかにする。

2-2. Laplace (1774)

BvM 定理の歴史を振り返る上で、その発見者として挙げられるのは Laplace であり、この点は学史研究においてほぼ同意が得られているようである。実際、彼が示した結果は極めて一般的な状況における事後分布の漸近正規性であるのみならず、事前分布の独立性や応用例をも含んだ顕著なものである。彼の示した結果は 1812 年初版の著作「確率の解析的理論」に含められる。この著作が後続の確率論や統計学の研究者らに多大な影響を与えたことを考えれば、Laplace の貢献は Bayes とは一線を画するものである。明確な事後分布の漸近正規性の初出は Laplace (1786) を待たねばならないが、そのおよそ 10 年前の 1774 年にその萌芽が見られる。

ここで取りあげるのは「事象の原因の確率の論文」と題された Laplace (1774) で、Bayes と独立に彼がベイズルールを提案した論文として知られるものである。論文は 7 節に分かれる：I. 序論、II. 原理の提示、III. 問題 I と解法（予測分布と事後分布の漸近理論）、IV. 問題 II と解法、V. 問題 III と解法（最適推定）、VI. 事前分布が考慮に入れられる問題、VII. 微分方程式の諸定理。特に、II 節において“原理”が提示され、それは一様事前分布を仮定したベイズルールと同等のものである。III 節において事後分布の漸近理論が展開され、この証明中に漸近正規性が現れる。

Laplace (1774) のモチベーションは原因の確率の決定法の提示とその応用である。彼は原因の確率の決定法として以下の原理を提示する：

原理：事象が k 個の異なる原因により生み出される場合、事象が与えられた下でこれらの原因の存在する確率はそれらの互いの関係が原因が与えられた下での事象の確率のようであり、それぞれの原因の存在する確率はその原因が与えられた下での事象の確率をそれぞれの原因が与えられた下での事象の確率の全ての合計で割ったものに等しくなる。(著者注：記法は原論文から変更した)

現在の記法で事象を x 、原因を $\theta_1, \dots, \theta_k$ とする。彼の原理は、事象 x を下にした原因 θ_i, θ_j の確率間の比とそれぞれの原因を下にした事象の確率の比が等しい、すなわち $P(\theta_i|x)/P(\theta_j|x) = P(x|\theta_i)/P(x|\theta_j)$ を意味し、特に

$$P(\theta_i|x) = \frac{P(x|\theta_i)}{\sum_{j=1}^k P(x|\theta_j)}$$

である。これは各原因の確率が、その原因の下での事象の確率と比例することを意味する。彼の原理は、逆確率 $P(\theta|x)$ が直接確率 $P(x|\theta)$ に比例することをもって、 θ の推論の原理として採用したものとと言える。言い換えれば、直接確率を用いて逆確率を計算することにより

推論を行う可能性を提示している。彼の原理がいかなる根拠の下で導出されたかは明瞭でないが、これは今日のベイズルールの一様事前確率の場合と同等のものである。ただし、事前確率の概念はこの原理には現れない。

続いて Laplace は以下の問題を提示した：“問題 I：壺に未知の比率で無限に多くの白と黒のチケットがある。そこから白を m 回、黒を $n - m$ 回の計 n 回引いたとき、新たにチケットを引いて白が出る確率を求める”。これは二項分布に基づく事後予測分布を求める問題に相当する。この問題に対する解答として、彼はまず白の出る未知確率 θ の事後密度として積分形の原理を提示する：

$$\frac{\theta^m(1-\theta)^{n-m}d\theta}{\int_0^1 \theta^m(1-\theta)^{n-m}d\theta}$$

これは Bayes (1764) によるベイズルールと同等のものである。その上で彼は事後予測確率を以下のように計算する： $\int_0^1 \theta^{m+1}(1-\theta)^{n-m}d\theta / \int_0^1 \theta^m(1-\theta)^{n-m}d\theta = (m+1)/(n+2)$ 。これはよく知られる Laplace の継起の法則である。この結果を踏まえ、さらに彼は、もし $m, n - m$ が大きければ白の出る確率が白の実際の相対頻度 m/n に近づくことを指摘しつつ、その結果を定理として示した。

定理 (Laplace, 1774)

$m, n - m$ が極めて大きいとき、不確実性が確実性にいくらかでも近づくようにして、壺に含まれる白のチケット数の全体のチケット数に対する比率 θ は、任意に与えた数より小さい数 $z > 0$ に対して $m/n - z, m/n + z$ の間に含まれる。

定理は現代の記法で $n \rightarrow \infty$ につれて $P(m/n - z \leq \theta \leq m/n + z) \rightarrow 1$ を意味し、 θ が相対頻度を中心とした非常に狭い区間 $m/n \pm z$ に含まれる事後確率が限りなく 1 に近づくことを意味する。すなわち白の出る確率は“ほぼ確実に”相対頻度に近い値であると推測できる。

事後分布の漸近正規性はこの定理の証明に垣間見える。実際、証明では、事後分布の分母と分子を今日の Stirling の公式などの漸近理論を用いて評価することで、以下で近似できることを示した：

$$\frac{1}{\sqrt{2\pi}} \int_{-z\sqrt{n^3/m(n-m)}}^{z\sqrt{n^3/m(n-m)}} e^{-t^2/2} dt.$$

ここで、 $n \rightarrow \infty$ につれて $n^3/m(n-m) \rightarrow \infty$ であり、今日 Gauss 積分として知られる $\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}$ を示しつつ、上式が 1 に収束することで定理は示される。注目すべきは、上式において θ の事後分布が $N(m/n, m(n-m)/n^3)$ で近似されることを意味する点である。すなわち、彼は $n \rightarrow \infty$ につれて θ が相対頻度 m/n にほぼ事後確率 1 で近づくことを示すのみならず、 θ の事後分布が今日の正規分布に従い、正確な分散まで導出しているのだ

ある。ただし、彼は正規分布について何ら言及しないことから、本人がその重要性に気付いていたかは判然としない。よって、これでもって事後分布の漸近正規性の導出と呼ぶことは難しいように思われる。それにも関わらず、実質的には事後分布の漸近正規性を含むことから BvM 定理の萌芽的成果と評価することはできよう。Hald (2010) によれば、1774 年当時彼は Bayes (1764) を知らなかったとされ、Bayes と独立の成果であるとされる。興味深いことに、Bayes と Laplace のいずれもがベイズルール の提案と同時に事後分布の漸近正規性の萌芽を導出しているのである。

注意として 1774 年の定理は事後分布の漸近正規性そのものを目標として示されたわけではない。この結果はあくまで、観測数が多いとき白の出る予測確率が相対頻度で近似されることを示そうとした中で暗に示されたものである。Laplace がこの結果に到達する動機はあくまで予測分布の近似にあったと見られ、この点で Bayes の近似公式とは動機が異なる。

2-3. Laplace (1786)

Laplace は 1774 年の論文からおよそ 10 年後、再び、原因の確率に関する論文を発表する。取りあげる論文は「極めて大きな数の関数である式の近似法の論文」と題された Laplace (1785) と翌年の「その続き」Laplace (1786) である。参照論文は再版だが、正確な初出年はそれぞれ 1782 年、1783 年である。構成は 4 部に分かれる：I. 高次べき乗を伴う変数をもつ微分可能な関数の近似的な積分 (1-7 節)、II. 有限および無限小の線形差分方程式の近似的な積分 (8-18)、III. これまでの方法の極めて大きな数をもつ様々な関数の近似への応用 (19-31)、IV. これまでの解析学の確率論への応用 (32-44)。1785 年論文は I、II、III 部を扱い、大きな数を含む関数の積分の近似理論と差分方程式や微分方程式への応用を含む。1786 年論文は IV 部を扱い、III 部までの数学の応用としてベイズルールに基づく確率論の諸結果が示される。特に IV 部 35 節で積分によるベイズルールと事前分布の仮定が提示される。36 節において事後分布の漸近展開が示され、特に漸近正規性を含む近似公式が示される。39 節以降では得られた近似公式の出生率や予測分布への応用例が示される。論文の内容は「確率の解析的理論」Laplace (1812) の基礎となるものであるが、そこに再掲されない結果も含まれる。

Laplace (1786) のモチベーションは、原因や未来事象の確率の決定法および大標本における近似公式とその応用に関連する。序論 (32 節) において彼は、決定論的な世界観の中でさえ生ずる偶然的現象は原因に対する人の無知に還元されることを指摘しつつ、確率とは原因の無知の表現に関連するものであると述べる。そして、単純な事象の確率が既知であれば複雑な事象の確率は組合せ論に帰着することを踏まえつつ、一方で次のように述べる：

偶然性の分析において最も興味深いのは、多くの場合、単純事象の確率が未

知である場合であり、過去の事象の中から未来の推測を導く手がかりを探すことである。(中略) この目的のために必要な原理と方法を別の場所で述べたが(著者注:Laplace (1774))、この方法は観測される事象の数が多いほどより正確になる利点がある。(中略) 私は新しくここに、多数の単純事象からなる結果の観察に基づき、これらの事象(中略) および未来の事象の確率を決定するための非常に便利な公式を提示する。

彼の問題意識は既知の観測から未知の原因の確率を求める逆問題にあり、1774年の自らの原理を踏まえたものであることを確認できる。そこから進んで、試行数が多い状況での公式について明言しており、実際これが漸近正規性の結果を含む。彼の“便利な公式の提示”というモチベーションは、純粋数学のみならず、統計的応用を見据えたものであることに注意したい。

Laplace は二項事象のような状況を想定しつつも、より一般的な状況を考えた。ある事象の未知の実現確率を θ とし、観測された事象の確率を θ の関数として一般に $f(\theta)$ と置く。ここで、 θ の異なる値それぞれを観測結果の“原因”と考え、 θ が θ_0, θ_1 ($0 \leq \theta_0 < \theta_1 \leq 1$) に含まれる事後確率を

$$\frac{\int_{\theta_0}^{\theta_1} f(\theta) d\theta}{\int_0^1 f(\theta) d\theta}$$

と提示する。ここで、 $f(\theta)$ は θ に関して微分可能な任意の関数を考えていることに注意したい。この関数は二項確率 $\theta^m(1-\theta)^{n-m}$ のみならず、大きな数 m_1, \dots, m_d に対してより一般的な $f_1(\theta)^{m_1} \dots f_d(\theta)^{m_d} \pi(\theta)$ の形式を許容する。特に、彼は強調こそしないものの、事後確率に微分可能な非一様事前分布 $\pi(\theta)$ が入ることを許容する。実際、彼は35節において

観測された結果とは独立に考えたときに、もし θ の各値が同等に可能でないならば、その確率を表す θ の関数を π と名付けると、直前の公式において f を $f\pi$ に変えれば十分である。このことは、 θ のすべての値が同等に可能であると仮定し、観測された結果は、確率がそれぞれ f と π である2つの独立な結果から形成されると考えることに帰着する。このようにしてすべての場合を、 θ の異なる値にすべて同等な可能性を仮定できる場合に帰着してよい。(著者注:記法は原論文から変更した)

慎重に読めば分かるように、彼は事前分布 $\pi(\theta)$ に一様分布を仮定すると言っているのではなく、 $f(\theta)\pi(\theta)$ により再び $f(\theta)$ と読み直せば、一般性を失わずに一様分布と仮定した場合と同じ形式に帰着できると言っている。実際、以下に示す彼の公式の証明においても $\pi(\theta)$ に非一様分布が許容されている。したがって、彼はここで非一様事前分布を許容した一般

的なベイズルールを提示した。

ここで、Laplace の公式を紹介する。 $f(\theta)$ の最大値を $\hat{\theta}$ とする。このとき、Laplace は θ の事後分布について以下の近似公式を示した。

公式 (Laplace, 1786)

未知 θ が $\hat{\theta}$ より小さな任意の定数 c 以下となる確率は、 $n \rightarrow \infty$ において、以下の公式で与えられる：

$$P(\theta \leq c) = \frac{1}{\sqrt{\pi}} \int_T^\infty e^{-t^2} dt - \frac{e^{-T^2}}{4\sqrt{\pi}} \left(-\frac{1}{2} \frac{d^2 \log f(\hat{\theta})}{d\theta^2} \right)^{-3/2} \frac{d^3 \log f(\hat{\theta})}{d\theta^3} + \dots$$

ただし、 $T^2 = \log f(\hat{\theta}) - \log f(c)$.

公式は n が大きいとき事後分布の近似の先導項が正規分布の定積分になることを意味し、事後分布の漸近正規性が明確に示されている。公式を提示した直後に彼は、“積分 $\int e^{-t^2} dt$ はこの分析で頻繁に遭遇するため、 $t = 0$ から $t = \infty$ までの積分の値の表を作成することは非常に有用であろう”と述べる。ここに当時まだ特別に名を付されていない正規分布の有用性に言及している。さらに、彼は右辺第二項において高次の展開まで評価していることも注目すべきである。

Laplace の公式では漸近分散が明確ではないが、独立同一分布に相当する $f(\theta)$ の設定では、漸近的に事後分布は最尤推定値を平均、Fisher 情報量の逆数を分散とする正規分布に従う。実際、 $f(\theta) = \pi(\theta) \prod_{i=1}^n f_i(\theta)$ という例を考え、 $J(\theta) = -\sum_{i=1}^n \frac{d^2 \log f_i(\theta)}{d\theta^2}$ とおく。もし $\hat{\theta}$ と c が近ければ、テーラー展開から近似的に $T^2 \approx (\hat{\theta} - c)^2 J(\hat{\theta})/2$ が従う。このとき $c = \hat{\theta} - z$ ($z > 0$)とすると、近似の先導項は以下で近似できる：

$$P(\hat{\theta} - z \leq \theta) \approx \frac{1}{\sqrt{2\pi J(\hat{\theta})^{-1}}} \int_{-z}^\infty e^{-J(\hat{\theta})t^2/2} dt.$$

これは θ の事後分布が $N(\hat{\theta}, J(\hat{\theta})^{-1})$ に従うことを意味する。ここで $\hat{\theta}$ は $\pi(\theta)$ を含めた $f(\theta)$ の最大値であり最尤推定値と同じではないが、 n が大きいとき $f(\theta)$ の中において $\prod_{i=1}^n f_i(\theta)$ が支配的になるため、適当な正則条件の下で $\hat{\theta}$ は最尤推定値と漸近同等である。したがって、この帰結は BvM 定理に相当する結果である。

さらに、 $J(\theta)$ は $\pi(\theta)$ に依存しないため、事前分布は 1 次の近似の意味で事後分布に影響しない。後述する von Mises と Bernstein は 20 世紀にやや異なる枠組みにおいてこの帰結を証明し、その貢献を強調することになるが、Laplace の結果はそれらに先んじることになる。Laplace はこの公式において事前分布に言及しなかったため、二人は Laplace の結果がそれを含むことに気付かなかったようである。Laplace の結果は事前分布の独立性

を数学的に含むが、言及がないのである。Laplace が非一様事前分布を扱い、事後分布の漸近正規性に事前分布が影響しない事実は Hald (2010) においても指摘されている。

Laplace の BvM 定理に関連する貢献の一つとして、その実証問題への適用が挙げられる。彼は 39 節において上記の近似公式を出生率の推測へ応用し、これは BvM 定理を用いた最初の実証例である。男子と女子の出生数をそれぞれ m 、 $n - m$ とし、男子の真の出生率を θ とする。1745 年から 1770 年のパリにおける観測値 $m = 251,527$ 、 $n - m = 241,945$ を用いて、男子の出生率が $1/2$ より高い事後確率を $1 - 10^{-42}$ 、つまり限りなく 1 に近いことを示す。これにより彼は“パリで観察された男子の出生数が女子を上回ることは、男子の出生の可能性が高いためであると考えて間違いないであろう”と結論付ける。より正規近似が意味を持つ小標本の例として、フランスのビトー村の観測値 $m = 203$ 、 $n - m = 212$ を用いて、女子の出生率が男子より高い事後確率を求める。正規近似に関する公式はこの例では以下に帰着する：

$$P\left(\theta \leq \frac{1}{2}\right) = \frac{1}{\sqrt{\pi}} \int_T^{\infty} e^{-t^2} dt + \frac{e^{-T^2\sqrt{2}}(2m - n)}{\sqrt{\pi mn(n - m)}}$$

ただし、 $T^2 = m \log m + (n - m) \log(n - m) - n \log(n/2)$ 。これを用いて、女子の出生率の方が男子より高い事後確率は 0.6702 と計算し、“この可能性はおよそ $2/3$ で示されているに過ぎず、相当数の出生が観測された全ての都市と同様に、ビトー村においても男子の出生率が女子よりも大きいと考える類推と整合をとるには、あまりに低すぎる”と結論付ける。ここにはある種の“検定”に近いアイデアが垣間見える。

この応用例は実証研究としても興味深いが、20万という大きな数のべき乗を含む多項式の積分計算の困難さを明確にする点でも重要である。観測数が極めて大きい場合の事後確率の計算は生来的に定積分の問題を含み、それまで知られた数学で直接的に解けなかったことがこの近似公式に結びついている。事後確率において漸近理論を考えることは、純粹数学的な興味のみならず、多数の観測値を直接扱う応用から自ずと要請された側面を持つことをここに確認できる。

Laplace の結果の歴史的重要性は、BvM 定理の発見とその解析的理論の提示にあると言える。1774 年の二項分布における萌芽的成果に続き、1786 年に彼は事後分布の漸近正規性を明確に発見した。彼はベイズルールが事前分布を含む形式で提示し、事前分布の事後分布への独立性をも本質的に示した。Hald (2010) によれば、Laplace は 1781 年には Bayes (1764) を知っていたとされるが、そうだとさえ、信用区間の導出を除き 1786 年の結果は Bayes の成果を顕著に改善したものである。さらに、彼は二項分布に留まらず、1 次元パラメータを伴う一般の滑らかな尤度関数に拡張して漸近正規性を示しており、この点は画期的である。彼は尤度関数の関数形が正規性の帰結に本質的でないことをこの時点で既に見抜いていたと思われ、これはまさに中心極限定理のベイジアンアナロジーである。ここでは詳しく取りあげないが、Laplace (1812) では自身の BvM 定理を再掲しつつ、さらに (de Moivre-) Laplace の定理 (邦訳 p.259) とその逆用法による信頼区間 (p.261)、

BvM 定理に基づく信用区間 (p.335)、2次元 BvM 定理の略証 (p.337) を含む。これらの成果を踏まえれば、疑いなく彼は BvM 定理の先駆者と評価できるであろう。加えて、彼による中心極限定理は確率論や統計学のみならず、その後の BvM 定理の進展にも影響をもたらすことになるが、この点を以降の項で明らかにする。

2-4. Bienaymé (1838)

Laplace により発見された BvM 現象について、次に進んだ結果を示したのは Bienaymé である。取りあげる論文は「観測値の平均の確率に関する論文 - Laplace ルールの直接的証明」と題された Bienaymé (1838) で、参考文献は再版だが初出は 1834 年である。彼の結果には新規性が認められるにも関わらず、Le Cam や Bernstein による文献には引用されず、どうやらその貢献は広くは認識されていないようである。von Mises (1919a) は Bienaymé の結果を引用し、Heyde and Seneta (1977) は BvM 定理のことを “Bienaymé-Mises result” と呼び、彼の貢献を強調した。論文は冒頭で確率論の逆問題と統計学への問題意識が述べられ、本論において二項分布、三項分布、そして多項分布における 1 次元パラメータの事後分布の漸近正規性が示される。結果それぞれについて今日の信用区間の構成がなされる。証明は現代の解析学から見れば厳密ではないが、詳細な数値分析により補完されている。

Bienaymé は数学的には多項分布のパラメータの事後分布の漸近正規性を示した。この点については疑いの余地はないようである。しかし、彼が純粋なベイズ問題の中で事後分布の漸近正規性に到達したかという点、実はそうではない。序論において彼は、二項事象の成功確率が既知の下で頻度が観測される確率を調べる問題だけでなく、観測された頻度を基に未知の成功確率を推測する逆問題が社会経済等の統計的研究で重要であることを指摘する。その逆問題へのアプローチとして、(1) ベイズルールを用いる方法、(2) de Moivre-Laplace の定理の逆用法の二つを挙げ、彼は後者がモチベーションであると明言する。すなわち、Bienaymé のモチベーションは純粋なベイズ問題にはないことになる。

ここでまず de Moivre-Laplace の定理の逆用法を説明する。ある二項事象の未知の成功確率を θ とし、 n 回の試行が行われ、成功が m 回観測されるとする。このとき de Moivre-Laplace の定理 (Laplace (1812)) から、直接確率で以下が従う：

$$P\left(\theta - z \sqrt{\frac{\theta(1-\theta)}{n}} \leq \frac{m}{n} \leq \theta + z \sqrt{\frac{\theta(1-\theta)}{n}}\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-t^2/2} dt.$$

ここで左辺を同値的に書き換えると、次を意味する：

$$P\left(\frac{m}{n} - z \sqrt{\frac{\theta(1-\theta)}{n}} \leq \theta \leq \frac{m}{n} + z \sqrt{\frac{\theta(1-\theta)}{n}}\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-t^2/2} dt.$$

このとき、もし θ に関する区間の境界に未知数が無ければ、 θ の信頼区間を構成できる。も

し境界に未知の θ がある場合でも、 θ を既知の m/n で置き換えることが正当化されるならば

$$P\left(\frac{m}{n} - z\sqrt{\frac{m(n-m)}{n^3}} \leq \theta \leq \frac{m}{n} + z\sqrt{\frac{m(n-m)}{n^3}}\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-t^2/2} dt.$$

を意味し、 θ の信頼区間 $m/n \pm z\sqrt{m(n-m)/n^3}$ を構成できる。ただし、 z は標準正規分布の分位点である。このような“置き換え”による信頼区間の構成法は Laplace (1812、邦訳 p.261) で最初に提示され、Bienaymé は“Laplace のルール”と呼んだ。また、直前の式は (de Moivre-) Laplace の定理における θ と m/n が全て逆転した形をとるため、“Laplace の定理の逆用法 (inverse)”とも呼ばれる (Todhunter (1865)、Keynes (1921) など)。ただし、上記の命題はあくまで直接確率による漸近正規性であることに注意したい。

実は、Laplace (1812) はこの置き換えによる区間の構成法を提示しつつも、直接確率におけるその正当性すなわち置き換え誤差が直接確率で漸近的に無視できることを厳密に証明していない。それどころか、置き換えの提示の直後の言及において唐突に“原因の確率から直接にこれらの結果に到達することもできる”と述べ、BvM 定理との混同が後に議論になったようである (この議論は Todhunter (1865) が詳しい。他に Czuber (1899)、Keynes (1921))。同書第 6 章には Laplace (1786) で得た二項分布に関する BvM 定理が再掲されるが、もちろんこれはベイズアンであり、直接確率の下での置き換えの厳密な証明を含まない。今日では、直接確率の下で置き換え誤差を無視できることは大数の法則と Slutsky の補題によって簡潔に正当化されている。しかし、1830 年代当時はこの事実についてまだよく分かっていなかったようである。

Bienaymé のモチベーションはまさにこの“置き換え”に関連する。彼の目的は置き換えの正当化にあると思われるが、それを直接確率の下で正当化することは目指さなかった。彼のモチベーションは、置き換えで得られる信頼区間と同じ境界を、“置き換えせずして” BvM 定理によって直接得られることを示すことだったようである。実際、序論後半において、彼は Laplace のルールの科学研究における重要性を説きつつ、Laplace のルールを使用することなく境界の定数 $\sqrt{\frac{m(n-m)}{n^3}}$ を見つける方法を確立することが論文の目的であると述べる。このために、彼はベイズルールから事後分布の漸近正規性を導くことで、その信用区間の境界に定数 $\sqrt{\frac{m(n-m)}{n^3}}$ が直接現れることを示した。すなわち、彼は Laplace の定理の逆用法の正当化を動機としながらも、その正当化にベイズアプローチを用いた。この点は非常に混乱を招く部分であると思われる。

Heyde and Seneta (1977) は Bienaymé の結果を“de Moivre-Laplace の定理の逆用法の厳密な証明へのベイズアプローチによる最初の例”との文言で評価した。しかし結果的には、彼の結果は置き換え誤差が漸近的に消えることのベイズ的正当化にはならないし、またそれを直接確率でも示していない。Bienaymé が示した命題は事後確率の漸近正規性

とそれを基にした信用区間の構成である。そうして、未知を既知で置き換えた際の境界の定数と同じものが、BvM 定理により置き換え操作を経ることなく直接的に得られることを示した。

ここで Bienaymé が示した新規的結果を紹介する。まず、二項事象が n 回試行され、 m_1 回成功、 m_2 回失敗するとし、成功と失敗の未知確率を θ_1, θ_2 ($m_1 + m_2 = n, \theta_1 + \theta_2 = 1$) とする。任意の既知定数 γ_1, γ_2 に対し、興味あるパラメータは θ_1, θ_2 の線形結合とする：

$$v = \gamma_1 \theta_1 + \gamma_2 \theta_2.$$

v の標本対応として $\mu = \gamma_1 m_1/n + \gamma_2 m_2/n$ とする。もし $\gamma_1 = 1, \gamma_2 = 0$ なら v は θ_1 を意味するため、 v の推測は θ_1, θ_2 を個別に推測する状況を含む。 v を推測対象とする根拠について Bienaymé は論文で明言していないようだが、これは θ_1, θ_2 の一般化というより、むしろ誤差論への応用を見据えたものと思われる。実際この設定を、もし成功すれば γ_1 、失敗すれば γ_2 が値として“観測”されると見なせば、 v は“観測値” γ_j の期待値を意味し、観測しようとする未知量の真値と見なせる。ただし、実際に彼の目的が誤差論にあったかどうかについては明瞭なことは言えないようである。

さて、動機に関する不明瞭な点を一度脇に置きつつ、Bienaymé の示した数学的命題を説明する。彼はベイズルールを明示しつつ、 v の事後確率として以下の命題を示した。

命題 (Bienaymé, 1838)

$n \rightarrow \infty$ につれて、 v の確率は

$$P\left(\mu - z \sqrt{\frac{m_1 m_2 (\gamma_1 - \gamma_2)^2}{n^3}} \leq v \leq \mu + z \sqrt{\frac{m_1 m_2 (\gamma_1 - \gamma_2)^2}{n^3}}\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-t^2/2} dt.$$

この結果に基づき、Bienaymé は v の信用区間を $\mu \pm z \sqrt{m_1 m_2 (\gamma_1 - \gamma_2)^2 / n^3}$ と提示した。ここで $\gamma_1 = 1, \gamma_2 = 0$ とすれば、 θ_1 の信用区間 $m_1/n \pm z \sqrt{m_1 m_2 / n^3}$ を得て、de Moivre-Laplace の定理の逆用法に基づく信頼区間と厳密に同じものをベイズから直接得たことを意味する。彼の結果はそれより一般的なものである。ただし前述のように、命題では置き換え誤差が無視できることを直接確率で示したわけではなく、またそれをベイズ的に示したものでもない。したがって、命題は信頼区間そのものの正当化にはなっていない。彼の結果はあくまで Laplace による置き換えから得られる $\sqrt{m_1 m_2 / n^3}$ と同一の定数を BvM 定理から直接得られるということである。

さらに、Bienaymé は多項分布に伴う 1 次元パラメータの結果を示した。 d 個の異なる事象について、それぞれの実現確率を $\theta_1, \dots, \theta_d$ とし、それぞれの事象を m_1, \dots, m_d 回観測するとする ($\sum_{j=1}^d \theta_j = 1, \sum_{j=1}^d m_j = n$)。既知定数 $\gamma_1, \dots, \gamma_d$ について、興味ある線形結合パラメータとその標本対応は以下である：

$$v = \sum_{j=1}^d \gamma_j \theta_j, \quad \mu = \sum_{j=1}^d \gamma_j m_j / n.$$

もし $\gamma_j = 1, \gamma_k = 0 (k \neq j)$ なら v は θ_j の推測を含む。ただし、この設定は多次元の推測は許容しない。もし d 個の事象それぞれについて、 $\gamma_1, \dots, \gamma_d$ が値として“観測”されると考えるならば、 v は観測値の期待値、 μ は観測値全体の標本平均である。ここで、改めて直接確率の立場から確率変数としての μ の標本分布を考えてみると、 θ_j, γ_j を固定的と見なすことで、

独立同一分布に関する Laplace (1810a, b) の中心極限定理は $\mu \sim N\left(v, \sum_{j=1}^d (\gamma_j - v)^2 \theta_j / n\right)$

という結果を含むことに注意したい。このとき、分散の未知 θ_j を既知 m_j/n で置き換えれば、

v の頻度論信頼区間は $\mu \pm z \sqrt{\sum_{j=1}^d (\gamma_j - \mu)^2 m_j / n^2}$ で与えられることが示唆される。

Bienaymé による明言はないものの、おそらく Laplace のこの結果を踏まえた上で、彼はベイズアプローチに基づき以下を示した：

命題 (Bienaymé, 1838, 多項分布)

$n \rightarrow \infty$ につれて、 v の確率は

$$P\left(\mu - z \sqrt{\frac{\sum_{j=1}^d (\gamma_j - \mu)^2 m_j}{n^2}} \leq v \leq \mu + z \sqrt{\frac{\sum_{j=1}^d (\gamma_j - \mu)^2 m_j}{n^2}}\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-t^2/2} dt.$$

結果から彼は v のベイズアン信用区間 $\mu \pm z \sqrt{\sum_{j=1}^d (\gamma_j - \mu)^2 m_j / n^2}$ を明示した。これは確かに、

Laplace の中心極限定理と置き換えに基づく v の信頼区間と同一の区間を、置き換えの議論を一切経由せずに BvM 定理から直接導出したことを意味する。すなわち、彼が序論で述べた目的をここに達成したことになる。ただし、やはり命題は信頼区間そのものの正当化にはなっていない。

まとめとして、彼は多項分布のパラメータの線形結合に関する新規的な BvM 定理を示した。彼のモチベーションは Laplace の定理の逆用法にあり、区間の未知を既知で置き換えることなくベイズアプローチから直接的に区間の境界を計算することにあつたようである。彼の本来のモチベーションは純粋なベイズ問題にはないようだが、結果として、ベイズアン信用区間の構成と置き換えに基づく信頼区間が一致することを証明した。モチベーションにおいて明瞭でない点が残るものの、数学的には新規的な結果を導出した貢献が認められる。ただし、彼の結果において事前分布は全く考慮されないことをここで指摘しておく。

Bienaymé の結果の歴史的な重要性は、信頼区間と信用区間の同一性と、それぞれの区間を互いに代替的に計算する可能性を示したことと言える。これは例えば信頼区間を計算す

ることが困難な頻度論推測において、計算された信用区間を信頼区間として代用できる可能性を示唆する。中心極限定理単独ではこの可能性を決して示唆しないため、BvM 定理は一見奇妙なこの方法の正当化に意義を持つことになる。特に、この代替の可能性について Laplace は明言していないことに注意したい。Laplace (1812) は中心極限定理の逆用法と BvM 定理の両方を示し、それぞれを基にした区間が結果的に一致する事実に関及したが (邦訳 p.261)、互いを代用する可能性を指摘しなかった。区間が一致することと、一方の区間をもう一方の区間で代用する可能性を指摘することには考え方に飛躍があると思われる。この意味で、Bienaymé が提起した区間の代用可能性は画期的なもので、現代において重宝される BvM 定理の重要な特徴の一つである (Chernozhukov and Hong (2003)、Kitagawa et al. (2020))。加えて、Heyde and Seneta (1977) は Bienaymé の意義について、区間の同一性により異なる手法間の違いが消え去ることを示唆する点に置いた。この点は重要であり、BvM 定理は異なる流派の推論間でロバストな結果をもたらす可能性を示唆する。Bienaymé によって示唆されたこれらの可能性は、ベイジアンの研究が直接的に頻度論の役に立ち、またその逆もあることを明確にする。

2-5. von Mises (1919a)

Bienaymé による結果以降の 19 世紀は、しばらくの間 BvM 定理に進展は現れなかったようである。この理由はおそらく 19 世紀初頭以降の確率論の興味が逆確率 (ベイズ推論) から直接確率すなわち中心極限定理、誤差論、最小二乗法などを中心としたものへ移行したことが考えられる。また、Cournot などによる主観確率や一様事前分布の批判も理由の一つに考えられる。Laplace ですら、1810 年周辺における中心極限定理の発見以降、彼の興味は逆確率から遠ざかったとされる (Hald (2010))。次に BvM 定理に進展が現れるのは 20 世紀に入ってからで、ほぼ同時期に、おそらく独立に von Mises と Bernstein によってなされる。

ここで取りあげる論文は「確率論の基本定理」と題された von Mises (1919a) である。同年に有名な“コレクティブ”に基づく確率の公理化に関する論文 von Mises (1919b) を公刊するが、1919a が先に書かれている。これら二論文は互いを引用しており、互いを補完する性格を持つようである。論文の内容の一部は、後に彼の書籍「確率計算とその統計及び理論物理学への応用」von Mises (1931) においても取りあげられる。確率の公理化や確率の哲学において 1919b が引用されることは多いが、BvM 定理を含む 1919a はそれほど広くは認知されていないようである。なお、1919a で扱われる中心極限定理は Fischer (2011) で詳細に取り上げられている。論文の構成は二部に分かれる：第 I 部 解析学的開発、第 II 部 確率論への応用。I 部は 6 節で構成され、II 部の基礎となる解析学と証明が示される。II 部は 4 節で構成され、コレクティブの定式化に基づく確率論と漸近理論、その応用が展開される (7. 二項数、ベルヌーイ、ポワソンの場合とその一般化、8. 第一基本定理、9. ベイズ問題とその一般化、10. 第二基本定理)。特に 8 節において中心極限定理、大数の

法則が取りあげられ、9節と10節において事前分布を許容した BvM 定理、多項分布に基づく多次元 BvM 定理、そしてより一般化された BvM 定理が取りあげられる。von Mises は確率の頻度説の創始者として知られるが、興味深いことに、彼は事前確率と事後確率をすべて頻度と捉えてベイズ問題を扱ったようである (von Mises (1964))。

von Mises (1919a) のモチベーションは序論冒頭に要約されている：

過去数十年の確率論の発展を追う者なら誰もが、数理科学のこの分野が 2 つの点で他の全ての分野に著しく遅れているという印象を受けずにいられないであろう。1 つには、ロシアの数学者による少数の研究を除き、確率論の解析的定理は他の解析学の分野では当然のこととなって久しい定式化と証明の正確さを欠いている。2 つめに、いくつかの価値あるアプローチにもかかわらず、数学分野としての確率論の基礎については未だにほとんど明確ではない。(中略) 本研究は前者の不足を補うために、確率論における一連の個別問題を統一かつ一般的な解析学の観点から取り扱うことを試みる。

彼は当時の確率論に関する二つの問題として (1) 確率論の解析的定理における定式化と証明の不正確さ、(2) 確率論の基礎が未確立であること、を指摘しつつ、1919a で前者に取り組むと述べる (後者は 1919b で扱われる)。これより彼のモチベーションは確率漸近理論の正確な定式化と証明にあることが明確になる。これは大数の法則、漸近正規性、誤差論などの諸定理について、直接確率のみならずベイジアンを含めて一つの枠組みで統一的に扱う試みである。1919a はこうした確率論の統合という大きな問題意識の下で諸結果が示されており、そこに一般化された BvM 定理が含まれる。

von Mises の BvM 定理のモチベーションは次の三点に関連すると思われる：(1) 数学的一般化、(2) 中心極限定理との対比、(3) コレクティブによる確率論の正当化。特に (2) は注目すべき点であり、彼は Laplace や Bienaymé の BvM 定理を一般化するのみならず、当時確率論における中心的テーマであった中心極限定理との対比も動機になっていたようである。実際、彼は中心極限定理と自身の BvM 定理を、それぞれ確率論の“第一基本定理”と“第二基本定理”と呼び、対照的かつ同格に扱う。より具体的には、直接確率とベイジアンそれぞれの漸近正規性、誤差論への応用、そして大数の法則を並列的に記述する。後の von Mises (1931, p.198-199) では直接確率とベイジアンの諸結果を表形式で対比し、互いのある種の双対的關係すら示唆した。この記述は現代の確率論体系には見られない極めて興味深いものである。彼は二群の諸結果の類似性を偶発的なものとは見なさず、相互の関連性を強調するのである。BvM 定理の一般化の方向性も、中心極限定理の形式と対応するように行っている。おそらく、このように二群の確率漸近理論の統合を強調することで (3) をも同時に目指したものと思われる。

von Mises が示した新規的結果として、まず d 個の値をとる多項分布の $d-1$ 次元パラメータに関する BvM 定理を紹介する。異なる d 種類の球が多く入った壺から n 個の球を取り出す問題を考える。それぞれの種類の球が取り出される確率を $\theta_1, \dots, \theta_d$ 、実際に取り出された個数を m_1, \dots, m_d とする($\sum_{j=1}^d \theta_j = 1, \sum_{j=1}^d m_j = n$)。 $\theta_1, \dots, \theta_{d-1}$ の同時事前密度は $\pi(\theta_1, \dots, \theta_{d-1}) = \pi_1(\theta_1) \dots \pi_{d-1}(\theta_{d-1})$ とする。このとき、取り出した結果が m_1, \dots, m_d である下での $\theta_1, \dots, \theta_{d-1}$ の同時事後密度は、ベイズルールに基づき

$$\frac{\pi(\theta_1, \dots, \theta_{d-1}) \theta_1^{m_1} \dots \theta_{d-1}^{m_{d-1}} (1 - \sum_{j=1}^{d-1} \theta_j)^{m_d}}{\int \dots \int \pi(\theta_1, \dots, \theta_{d-1}) \theta_1^{m_1} \dots \theta_{d-1}^{m_{d-1}} (1 - \sum_{j=1}^{d-1} \theta_j)^{m_d} d\theta_1 \dots d\theta_{d-1}}$$

ただし、分母の積分領域は $\sum_{j=1}^d \theta_j = 1$ を満たす範囲である。

定理 (von Mises, 1919, 多項分布)

異なる d 種類の球の入る壺から極めて多くの球を取り出すとき、それぞれの種類の球を一度取り出す確率 $\theta_1, \dots, \theta_{d-1}$ の同時事後密度は、事前密度にかかわらず、次の $d-1$ 変量正規分布に従う：

$$P\left(\theta_1 \leq \frac{m_1}{n} + \frac{z_1}{\sqrt{n}}, \dots, \theta_{d-1} \leq \frac{m_{d-1}}{n} + \frac{z_{d-1}}{\sqrt{n}}\right) \\ \rightarrow \frac{1}{\sqrt{\pi^{d-1} m_1 \dots m_d / n^d}} \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_{d-1}} e^{-Q(\theta_1, \dots, \theta_{d-1})/2} d\theta_1 \dots d\theta_{d-1},$$

ただし、 $Q(\theta_1, \dots, \theta_{d-1}) = n \sum_{j=1}^{d-1} \theta_j^2 / m_j + n (\sum_{j=1}^{d-1} \theta_j)^2 / m_d$.

定理は多次元パラメータの同時分布での漸近正規性を意味し、Laplace と Bienaymé の結果を一般化した新規的なものである。さらに結果は事前分布が漸近分布に依存しないことを意味し、彼はその点を明確に強調する。正規分布に従う変数の線形写像は再び正規分布である事実に注意すれば、この定理は Bienaymé の多項分布の結果を含むことが分かる。

次に、von Mises による最も一般化された結果として“第二基本定理”を紹介する。この記述には、コレクティブの定義やそれに伴ういくつかの特別な概念が必要になる。特にコレクティブは確率変数という概念を持たず、その記述は現在の Kolmogorov のものと極めて異なる。したがって、それらをここで詳しく記述するのはあまり有益ではない。ところが実は、彼の設定は現代の言葉に翻訳すれば標準的な正則条件と本質的に近いものである。そこで、やや不完全な記述になることを断った上で、以下では現代の言葉に翻訳した結果を紹介する。

d 次元パラメータベクトルを θ とし、パラメータ空間を R^d の部分集合 Θ とする。大きさ n の独立な観測値を x_1, \dots, x_n とし、 θ の下で x_i が実現する確率を $f_i(x_i|\theta)$ とする。この設定は観測値の非同分布を許容する。 $f_i(x_i|\theta)$ に対する正則条件として、 θ に関し一意の最大値を持ち、かつ θ に関して3回連続微分可能で全ての導関数が下からも上からも一様に有界と仮定する。 $f_i(x_i|\theta)$ の θ に関する最大値を $\bar{\theta}_i$ とし、 $f_i(x_i|\bar{\theta}_i) = 1$ と仮定する。 θ の同時事前密度

$\pi(\theta)$ に対し同様の滑らかさの正則条件を仮定する。ベイズルールから θ の事後密度は以下である：

$$\frac{\pi(\theta) \prod_{i=1}^n f_i(x_i|\theta)}{\int_{\Theta} \pi(\theta) \prod_{i=1}^n f_i(x_i|\theta) d\theta}$$

ここで $J_i = -\frac{\partial^2 \log f_i(x_i|\bar{\theta}_i)}{\partial \theta \partial \theta^T}$ とし、 $\bar{J} = \frac{1}{n} \sum_{i=1}^n J_i$ を可逆とする。また $\bar{\theta} = \bar{J}^{-1} \frac{1}{n} \sum_{i=1}^n J_i \bar{\theta}_i$ とする。

定理 (von Mises, 1919, “確率論の第二基本定理”)

すべての i に対し、 $f_i(x_i|\theta)$ は正則条件を満たすとす。このとき $n \rightarrow \infty$ につれて、 d 次元パラメータ θ の事後分布は、 $\bar{\theta}$ を平均、 $(n\bar{J})^{-1}$ を分散とする d 変量正規分布に従う。

定理では事前密度と尤度関数はともにパラメトリックに特定化されておらず、正則条件を満たす任意の関数が許容されている。von Mises の実際の第二基本定理はおそらく変数の正規直交化を用いることで漸近分散を記述しないが、簡単な計算から本質的に $(n\bar{J})^{-1}$ であることを確認できる。定理は標本の非同一性を許容する一般的な命題だが、 $\bar{\theta}$ の解釈は難しい。定理は同一分布の場合も同様に証明でき、実際 $\prod_{i=1}^n f(x_i|\theta)$ の最大値を $\hat{\theta}$ とすると、

事後分布は $N(\hat{\theta}, (nJ(\hat{\theta}))^{-1})$, $J(\hat{\theta}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(x_i|\hat{\theta})}{\partial \theta \partial \theta^T}$ である。情報等式を仮定すれば、

このとき事後分布の平均は最尤推定値、分散共分散行列は Fisher 情報行列の逆行列と漸近同等であり、結果は今日知られる独立同一分布における BvM 定理のバージョンである。ただし、彼はこの定理に関し θ の信用集合 (区間) の構成については議論していない。

第二基本定理を誤差論へ適用した例を紹介する。彼は計測したい未知量について、その観測値に基づきベイズ推測を行うという誤差論を示した。ここで、ある未知量の真値 x を知るために n 回の観測を行うとする。観測ごとに d 種類の観測値 y_1, \dots, y_d のどれかが得られるとし、それぞれの得られる確率を $\theta_1, \dots, \theta_d$ 、実際に観測された頻度を $m_1/n, \dots, m_d/n$ とする。このとき、未知量の真値の事後確率について、以下の定理に従う。

定理 (von Mises, 1919, “誤差論の第二定理”)

観測値の数 n が大きければ、未知量の真値 x の事後密度 $P(x)$ は、事前密度にかかわらず、次の正規分布に従う：

$$P(x) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{(x-\mu)^2}{2s^2}}, \quad \text{ただし} \quad \mu = \sum_{j=1}^d y_j m_j / n, \quad s^2 = \sum_{j=1}^d (y_j - \mu)^2 m_j / n.$$

結果として事後平均と分散は観測値全体の標本平均と標本分散に対応する。興味深いことに、この結果は Bienaymé の多項分布の命題と形式上同等のものである。また、この定理

は第二基本定理のみによって示されることに注意したい。中心極限定理はどこにも使われず、また観測誤差の生成分布として正規性は仮定されていない。すなわち、誤差の生成分布が特定されなくとも、観測値に基づく未知量の事後分布は漸近正規することを示している。言い換えれば、漸近的な事後分布としては、通常誤差論で用いられる誤差の正規性は本質ではないことがここで明らかになる。この応用において、第二基本定理が許容する尤度関数の任意性の意義が明確になる。さらに、結果は事前分布の選択にも依存しないため、第二基本定理における事前分布からの独立性が重要な役割を果たしている。

まとめとして、von Mises は多次元パラメータに関する一般的な BvM 定理を示し、その意義を明確に強調した。彼の BvM 定理のモチベーションは直接確率を含めた確率論の定式化と漸近理論の中にあり、数学的一般化や中心極限定理との対比と関連する。彼の結果は標本の非同一性、多次元、滑らかな尤度関数、非一様事前分布を許容した顕著な一般性を持つ。その誤差論への応用は定理の特性が見事に示された例である。ただし、統計的応用への意識は高くはなかったようで、信用区間の構成には全く言及せず、信用区間と信頼区間の同一性に関する指摘も見られない。

von Mises の結果の歴史的な重要性は、BvM 定理の数学的一般化の一つの完成であると言える。彼の結果は非同一分布、多次元、滑らかな尤度、非一様事前分布を許容した一般性を持ち、まさに多変量中心極限定理のベイジアンアナロジーと言える。特に、彼は BvM 定理を独自の確率の公理の枠組みにおいて示し、中心極限定理との双対的性格を認識し、そこに BvM 定理の重要性を置いた。それは彼の最も一般化された結果に対する“第二基本定理”という名にも現れている。このような態度は Laplace (1812) には見られなかった。とりわけ、統計学というより確率論的興味においてであろうが、彼は事前分布の漸近分布への独立性を強調した。彼は事前分布の特定化が漸近的にベイズ推論において本質的でないことを示している。Laplace に始まる BvM 定理の数学的一般化は、有限次元パラメータをもつ滑らかなモデルの枠組みにおいて彼によって一度完成したと言ってよい。ただし、後述する同時期の Bernstein の貢献の全てを含むわけではない。von Mises は統計学への問題意識は高くなかったようで、信用区間の構成やその信頼区間との関連等の結果を含まない。彼の一般化志向は統計的応用から直接要請されたものというより理論的なものようである。この点は Bienaymé や Bernstein と大きく異なる点である。

2-6. Bernstein (1927, 1934)

ここで取りあげる文献は「確率論」と題された教科書 Bernstein (1927) である。本書の初版は 1927 年であり、1934 年、1946 年に改訂され第 4 版まで出版された。Videnskii (2000) によれば本書は当時のロシアの学生に広く読まれたと評されている。本書の構成は 5 部に分かれる：第 1 部. 確率論の基礎、第 2 部. 数学的確率の主要な計算法、第 3 部. 大数の法則、第 4 部. 正規確率の法則、第 5 部. 分布曲線と曲面の理論の基礎。特に第 1 部で Bernstein による確率の公理が提示され、第 3 部、第 4 部において確率の漸近理論とその

応用が提示される。

内容に入る前に、まず Bernstein による BvM 定理の初出文献については注意が必要である。同書は 1917 年が初版としてしばしば引用されるが (Le Cam (1953)、Le Cam and Yang (1990)、Lehmann and Casella (1998)、Heyde and Seneta (1977))、正しくは 1927 年が初版だと思われる (この点は Seneta (2013) も示唆している)。この誤りが生まれた原因はおそらく、同書 p.271 において BvM 定理を“1917 年のハリコフ大学 (著者注: 当時の所属機関) の自身の講義ノートで示した”と言及したことによると思われる。おそらくこの言及により、彼の BvM 定理は 1917 年が初出と見なされ、それがさらに一部の研究者に同書の初版と誤解された可能性がある。しかしながら、このノートは公刊されたか不明の著作物であり、実際、彼の書誌情報 (Forty years of mathematics in the USSR, Vol.II, P.71) によれば、1917 年に確率論の著作 (後述する確率の公理に関する論文を除き) を出版した記録はない。その後、第 2 版 Bernstein (1934) に“1917 年の証明を第 4 補論で再掲する”と追記され、ここによりややく証明を確認することができる。1934 年版に 1917 年の証明をあえて再録したことは、裏を返せば、参照先のノートが 1934 年時点で入手不可か、あるいは公刊されていない可能性を示唆する。以上の考察から、本論文では明らかな公刊が認められ、かつ BvM 定理への言及と応用を含む Bernstein (1927) を初出と仮定する。いずれにせよ、確かな公刊物として完全な証明が示されるのは、1927 年ですらなく、1934 年である。

Bernstein が同書で目指したことは、自身の公理に基づく確率論の定式化、漸近理論とその応用の提示である。前書きにおいて彼は、“本書の目的は物理学、生物学、経済学などへの多彩な応用の基礎となる確率論の原理と一般的な方法を提示すること”と述べる。特に、第 1 部第 1 章で確率論の動機を述べつつ、自身による確率の三公理を導入する。この公理の初出は (紛らわしいが) Bernstein (1917) の論文で、これは歴史上最初の確率の公理と評され (Maistrov (1967))、第 1 部第 1 章と第 2 章はそれを再掲したものである。この公理から加法定理、乗法定理、ベイズルールが導かれ、特に第 3 部と第 4 部において大数の法則と中心極限定理を含む漸近理論の諸結果が整理される。この中で彼はベイズ問題と統計的推測を取りあげ、特に第 4 部第 4 章の標本調査の中で BvM 定理が示される。

Bernstein の BvM 定理のモチベーションは統計的推測、特に Laplace の定理の逆用法と信頼区間の正当化に関連したものである。すなわち、彼は純粋なベイズ問題の一般化として BvM 定理に到達したわけではない。ところが、彼のモチベーションが Bienaymé と同じものであるかという点、そうでもなさそうである。実は、信頼区間の文脈において Bernstein が BvM 定理を示した論理は明瞭でなく、彼が何をしようとしていたのかは極めて曖昧なのである。

これを見るため、以下に第 4 部第 4 章“統計学における標本調査”の概要を説明する。第 1 節冒頭において、Bernstein は十分大きな数の対象を持つ母集団がある特徴 A をもつ真の比率 θ を推測する問題を提示する。未知 θ を推測するため、母集団から大きさ n の標本を

取り出し、特徴 A を持つ数を m とする。ここで、彼は (de Moivre-) Laplace の定理とその逆用法を述べつつ、 θ の信頼区間 $\frac{m}{n} \pm z \sqrt{\frac{m(n-m)}{n^3}}$ を提示する。その直後、彼は唐突に

この公式の完全な正当化には、ベイズルールに基づく Laplace の定理の逆 (inverse) を直接的に証明する必要がある。

と述べ、その脚注において非一様事前分布を許容したベイズルールに基づき BvM 定理の略証を示す。1927 年版では略証のみが示されるが、形式的な定理と証明は 1934 年第 2 版第 4 補論で改めて示される。

定理 (Bernstein, 1927, 1934, “Laplace の定理の逆”)

もし特徴 A の実現が一定の確率 θ を持ち、不等式 $\theta_0 < \theta < \theta_1$ の事前確率が連続関数 $\pi(\theta) \geq 0$ について $\int_{\theta_0}^{\theta_1} \pi(\theta) d\theta$ なら、 n 回の実験で特徴 A が m 回起こる場合、以下の不等式の確率は、 $n \rightarrow \infty$ につれて

$$P\left(z \sqrt{\frac{m(n-m)}{n^3}} < \theta - \frac{m}{n} < z \sqrt{\frac{m(n-m)}{n^3}}\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-t^2/2} dt.$$

ただし、ある $\lambda > 0$ について $\pi(m/n) > \lambda > 0$ 、ある $\epsilon > 0$ について $1 - \epsilon > m/n > \epsilon$ とする。

定理と証明は疑いなくベイズルールに基づくもので、信頼区間と同一の境界を持つ区間に関して、 θ の事後分布が漸近正規することを示している。また、明示的に連続な事前分布 $\pi(\theta)$ が考慮に入れられ、事前分布が漸近分布に依存しないことを示し、彼は事前分布の独立性を明言する。今日から見て、結果は $\frac{m}{n} \pm z \sqrt{\frac{m(n-m)}{n^3}}$ がベイジアン信用区間であることを意味し、その意味で妥当なものである。しかし、これは直接確率の下で置き換え誤差が無視できる証明ではないため、直接確率における信頼区間の正当化としては妥当ではない。どうやら、この結果は彼が本来目的としたはずの問題の解答にはなっていないように思われる。さらに不思議なことに、第 1 節の残りにおいて Laplace の置き換えに伴う誤差の議論に戻り、小標本における置き換え誤差を数値例から検討する。さらに、続く第 2 節から第 5 節では、第 1 節の結果を基に非復元抽出の場合の置き換え誤差、比率の差の推測、そして母平均の推測が扱われるが、これらの節でも中心極限定理と置き換えを基にした区間とそれに対するベイジアンの正当化が何度も行われる。

Bernstein が BvM 定理を示した論理については、慎重に検討する必要があるだろう。標本調査を目的に Laplace の定理と逆用法による信頼区間を提示した一方で、しかしその正当化には直接確率の論理ではなく唐突にベイジアンの論理を持ち出す。しかも、彼はそれを信

頼区間の“完全な正当化”と主張した。この議論は第4章第2節から第5節も同様で、中心極限定理を基にした置き換えとその正当化にベイズアプローチを用いる議論が続く。ここに彼の定理の分かりにくさが存在する。

Bernstein はいかなる論理で BvM 定理を置き換えるの正当化とみなしたのかという点について、以下で検討を試みる。まず、彼の状況は Bienaymé と近いように見えるが、定理を示す動機はどうやら同じではない。もし動機が同じなら定理を示した後の置き換え誤差の議論は不要であることに加え、何より彼は信頼区間の境界を“置き換えなしに” BvM 定理から計算できるということに一切言及しない。一つの可能性として、彼は直接確率と事後確率を同一視（あるいは混同）した上で、定理により置き換え誤差を無視できると考えていた可能性が挙げられる。もしそうなら、置き換えた区間の事後確率の意味での漸近正規は、直接確率の意味での正規分布とも見なせることを意味する。このとき置き換え誤差が消えた形で正規分布に収束したかのように見えるため、BvM 定理により置き換え誤差を無視できると信じたとしてもおかしくなく、彼が“完全な正当化”と主張したことに説明がつく。もちろん、Bernstein ほどの著名な数学者が直接確率と事後確率を同一視あるいは混同した可能性はさすがに憶測の域を出ない。しかし、第4章全体における Laplace の定理の逆用法と BvM 定理をあたかも同一視するかのよう一連の議論からすれば、あり得ないことではない。また、同書は独自の公理から確率を定義するため、書籍全体を通して直接確率と事後確率を区別していない可能性がある。実際、彼は第2部第3章1節における事後確率の定義に際して、それを数学的条件付き確率と区別しないと明言する。もう一つの根拠として、彼の定理は疑いなくベイズルールに基づく BvM 定理にも関わらず、本人は“Laplace の定理の逆 (inverse)”と呼んでいる事実である。ところが Laplace の定理の逆と呼ぶときは、典型的には逆用法すなわち Laplace の定理で得られた区間の未知を既知で置き換えたものの直接確率での漸近正規性を意味するはずである (Todhunter (1865)、Keynes (1921))。Bernstein の貢献を整理した Seneta (2013) は、彼の定理のモチベーションについて言及していない。いずれにせよ、彼の定理は信頼区間の正当化にはならず、 $\frac{m}{n} \pm z \sqrt{\frac{m(n-m)}{n^3}}$ が信用区間として妥当なことを示しただけのように見える。

さて、Bernstein は第4章後半において BvM 定理に基づくいくつかの統計的応用例を示しているのは、von Mises と異なる態度が示された興味深い点である。一例として、独立な2期間において労働組合の女性比率がそれぞれ0.215, 0.2007で、標本の大きさをそれぞれ $n_1 = 1,000$, $n_2 = 1,500$ とする。このとき、標準偏差はそれぞれ1.29, 1.034で、比率の差の標準偏差は $\sqrt{1.29^2 + 1.034^2} = 1.31$ と計算する。比率の差1.43と標準偏差1.31を比較し、彼は“女性比率が本当に下がったと信じる深刻な理由はない”と結論した。さらに、同じ比率がさらに大きなデータで得られた場合として $n_1 = n_2 = 10,000$ を考え、比率の差の標準偏差0.57から $1.43/0.57 = 2.5$ と計算した。この値が標準正規分布において出現する確率はわずか約1.2%とした上で、彼は“女性比率の紛れもない減少を示すであろう”と結論付

けた。これらの解釈において“検定”のアイデアが垣間見えることが分かる。ただし、上記で用いられた比率の差の標準偏差の公式に関する BvM 定理による証明はなされていない。さらに、続く節では母平均の推定問題において BvM 定理の結果らしきものが見られる。しかしベイズルールに基づく証明を含まないことに加え、やはり逆用法と混同して示された可能性を否定できず、新規的な BvM 定理とは見なせないようである。

Bernstein の結果の歴史的重要性はむしろ、事前分布の特定化が漸近的には信用区間の構成に本質的でないことを示した点にあると指摘できる。これは信用区間が事前分布の特定化に関しロバストであることを意味する。彼がいかなる論理で信頼区間の正当化のために BvM 定理を示したかは明瞭でないものの、結果的に、彼の定理は信用区間と信頼区間の一致が事前分布の特定化に依らずロバストに成立することを示唆する。信頼区間を事前分布に依ることなく信用区間で代用できる可能性には重要な意義があり、この点は Bienaymé の結果には含まれない。また、Bernstein は確率論の統計的応用を重視し、実行可能な信頼区間（信用区間）の正当化にこだわった点は、von Mises と異なる態度である。これらの点において、Bernstein 独自の貢献が認められる。ただし、Bernstein の結果は Laplace、Bienaymé から顕著な一般化があったと言えるかは疑問が残る。彼の結果は尤度関数や次元に関して一般性が高いと言えず、これは同時期に von Mises によって達成された顕著な一般化と対照的である。Seneta (2013) は彼の結果について“(Bernstein の) この定理のバージョンは現在 Bernstein-von Mises の定理と呼ばれているが、この帰属はあまり適切ではない”と評価した。また、Bernstein の結果の初出は 1917 年というより 1927 年と見る方が妥当と思われ、von Mises (1919a) に先行するというには情報が不十分なようである。いずれにせよ、Bernstein の貢献が von Mises と同格のものかはおそらく議論の余地があると思われる。

3. 現代的視点からの古典的 BvM 定理の意義

ここまで古典的な定理の諸結果について振り返った。歴史的な背景を振り返ると、BvM 定理の進展はベイズ問題の数学的一般化の一途を辿ったものとは限らず、そのモチベーションは貢献者の間で大きく異なることが分かる。特に、以下の二点は重要な知見であると思われる：(1) 中心極限定理の発展が BvM 定理の進展の契機や刺激になったこと、(2) 20 世紀の BvM 定理の背景に確率の公理化の運動が影響したこと。

第一に、BvM 定理の発展において指摘すべき点は、Laplace を始めとする中心極限定理の発展がその契機や刺激になったことである。当初の BvM 定理は逆確率あるいはベイズ問題に対する問題意識の下、予測分布や近似公式を導出する中で発見された。ところが 1810 年周辺を中心極限定理とその逆用法の提示以降、おそらく研究者らの問題意識が直接確率を中心としたものに移ったことで BvM 定理の研究は発見後、間もなく停滞したようである。Gauss (1809) は誤差論においてベイズアプローチを考えつつも事後分布の漸近正

規性には興味を示さなかった。Poisson (1830、1837) は Laplace の BvM 定理を出生率等の応用に適用するが、自らそれを一般化させることはしなかった。そうした中、Bienaymé や Bernstein は直接確率の信頼区間の正当化に動機付けられ結果的に新たな BvM 定理を導出した。興味深いことに、BvM 定理のいくつかは中心極限定理に直接影響を受けて示されたのである。また、von Mises は 19 世紀後半以降の Markov や Chebyshev らによる中心極限定理の発展に刺激されたようで、中心極限定理を第一基本定理と呼び、そのベイジアンアナロジーとしての自身の BvM 定理を第二基本定理と呼んだ。中心極限定理の発展なくして定理のこうした発展過程を説明することは難しいであろう。現在のベイジアンと頻度論は哲学的に別の流派と見なされるが、古典的 BvM 定理の発展はこれらの相互の関わりの中にあったことを指摘できる。この点は定理に至る動機を慎重に検証することで明確になった点である。

第二に、von Mises と Bernstein の結果の背景に確率の公理化が影響した点を指摘できる。20 世紀初頭において確率論が数学の一分野としてまだ明確になっていないとき、確率の公理化の運動が起こった（公理化の歴史は Maistrov (1967) が詳しい）。この動きに対する最初の提案が Bernstein (1917) (Bernstein (1964) に再録) とされ、そのすぐ後に von Mises (1919b) において提示される。二人の確率の公理は全く別の体系をとり、Bernstein は代数的構造を持つのに対し、von Mises は仮想的な実験の観測列であるコレクティブによる頻度の極限に基づく。公理化により、確率は哲学から切り離され、ベイズルールは公理のみから導出された完全な数学的対象となった。さらに、公理化により大数の法則や中心極限定理の定式化を明確にする漸近理論の土台を獲得した。こうした確率論の再構築が von Mises (1919a)、Bernstein (1927) (あるいは 1917 年のノート) の基礎になり、ベイズルールの漸近挙動を再び研究する契機になったと考えられる。その後、結果として集合と測度に基づく Kolmogorov の公理が残るが、当時の公理化運動は二人の結果の導出に一定の役割を果たしたと思われ、Bienaymé からしばらく後になって二人が同時期に、おそらく独立的に定理を示すという特異な現象を生み出した。20 世紀初頭の BvM 定理の成立においてこの点は重要だと思われるが、二人の貢献への言及を含む既存の学史文献では指摘されていない。

ここで古典的諸結果の概観を踏まえ、現代の BvM 定理との関連性や相違点を明らかにする。現代の BvM 定理として重要な次の三点についてその起源と最近の展開との関連を見ていく：(1) 尤度関数や事前分布の特定化に依存しない漸近正規性、(2) ベイジアンと頻度論の漸近分散の一致あるいは信用/信頼区間の一致、(3) 無限次元パラメータを含む問題への拡張。

まず、事後分布の漸近正規性に関し現代から見て重要な点は、尤度関数の任意性である。Laplace (1786) は二項分布を一般化した滑らかな尤度関数を許容した最初の結果である。1774 年の漸近正規性の萌芽的成果の後、彼はその帰結に二項分布という関数形が本質的でないことに気付いたものと思われる。その結果は 1786 年に結実するが、この点は疑いな

く Laplace の顕著な貢献と言える。後に von Mises は次元を一般化し、多次元パラメータを伴う滑らかな尤度関数の下で事後分布が多変量正規に収束することを示した。彼らによる尤度関数の一般化は現代において重要なものである。含意する一例として、これはデータ生成過程の特定化にロバストな推論の考えに繋がる。例えば真の分布が正規分布でない場合においても、正規分布を想定した尤度に基づく推論が正当化される根拠になる。一般化の観点からは、事前分布の事後分布への独立性も数学的には Laplace に遡ることができ、これは von Mises と Bernstein に一世紀以上先んじる。すなわち、観測数が多い場合のベイズ推論には事前分布は本質的な役割を果たさないことがベイズ統計学史の極めて初期から認識されていたことになる。事前分布の特定化については後に Fisher らによって批判を受けるが、ベイズ推論が再評価された現代においてこの性質は事前分布の議論を緩和させる根拠になる。Diaconis and Freedman (1986) は、観測値が多ければ事前分布に対し異なる主観的信念を持つ主観的ベイジアンの間でさえ共通の結果に到達しうることを指摘した。

次に重要な点はベイジアンと頻度論の漸近分散の一致、特に信用区間と信頼区間の一致についてである。古典的定理は最尤法や Fisher 情報量が提示される以前に導出された結果であるため、最尤推定量の標本分布との対比は明確に認識されていないと考えられる。それにも関わらず、二項分布の文脈において、信用/信頼区間の一致は Laplace (1812) に短く言及があり、区間の代用可能性については Bienaymé (1838) が示唆したと言える。これより、二つの区間の相互の関連性は 19 世紀前半には認識されていたことを確認できる。事前分布の事後分布への独立性を明言した von Mises と Bernstein の結果は、二つの区間の関連をさらに確かなものにする。相異なる方法間で同一の結論に到達しうる可能性は、現代から見て疑いなく重要な帰結である。また、代用可能性の提起は現代において重要な BvM 定理の特徴の一つであり、最近ではセミパラメトリックモデルに関する Chernozhukov and Hong (2003)、Norets (2015) など、マルコフ連鎖モンテカルロ法とブートストラップの間の代用に関する Kitagawa et al. (2020) がある。

最後に、現代では無限次元パラメータを含む BvM 定理の研究が進んでいるが、古典的諸結果にこうした問題意識はなく、母集団分布の密度推定すら考えられていない。このことから、無限次元のパラメータに関する問題は現代における新しい問題意識といえる。実際、20 世紀後半の Freedman らによる研究により、無限次元パラメータの推測においてベイジアンと頻度論で結果が一致するとは限らない点が明らかにされた。ここから、信用/信頼集合の一致はパラメータが少ない場合に特有の結果だった可能性が示唆され、これはベイジアン-頻度論間の基本的な違いを改めて浮き彫りにするものである。一方で、最近では一部のノンパラメトリックモデルやセミパラメトリックモデルにおいて再び結果の一致が見られる研究成果が数多く報告されている。特に、ノンパラメトリックモデルでは信頼集合との厳密な一致というより、信用集合の頻度論被覆確率に直接的に焦点を当てる研究も多く存在する (Castillo and Nickl (2013) など)。次元が高い場合における双方の統計学の関連性は、今後のさらなる研究において明らかにされるものと期待される。

4. 結論

古典的な BvM 定理の諸結果を概観することで本論文が明らかにしたことは以下に集約される：(1) ベイズ推論の歴史の最初期から漸近正規性は発見された。(2) 古典的結果に至る動機は貢献者の間で大きく異なり、とりわけ定理名に帰属する Bernstein と von Mises ですらその動機は異なる。(3) 古典的結果の発展には中心極限定理の成立と発展が強く影響し、一部において頻度論信頼区間の正当化と直接的に関連する。(4) 20 世紀の Bernstein と von Mises の結果の成立の背景に確率の公理化が影響した。(5) これまでの学史で過小評価された Bayes の近似公式は BvM 定理の萌芽であり、正確な信用区間を含む。これらの知見は既存文献で指摘されておらず、BvM 定理に焦点を当てて概観したことで明確になった。

古典的な BvM 定理の諸結果においてはベイジアンと頻度論が交錯する場面がいくつも見られ、ここに BvM 定理の研究の意義があるように思われる。定理は純粋なベイジアンの帰結というより頻度論や中心極限定理との相互の刺激によって発展し、それは現在も続いている。こうした知見はベイジアンと頻度論のどちらか片方を考えるだけでは分からない部分であり、BvM 定理を通してベイジアンと頻度論の相互の可能性をより深く理解できる可能性がある。双方の考え方は基本的な哲学から異なるが、深いところで部分的に結びつき普遍的な共有物を持つ側面を否定できないようである。今後の BvM 定理と双方の統計学のさらなる研究の進展により統計的推論の全体像が明らかになることが期待される。

引用文献

- Bayes. T. (1764). "An Essay towards solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. communicated by Mr. Price, in a letter to John Canton, A.M. F.R.S.," *Philosophical Transactions*, 53, 370-418.
- Bayes. T. (1765). "A Demonstration of the Second Rule in the Essay towards the Solution of a Problem in the Doctrine of Chances, . published in the Philosophical Transactions, Vol. LIII. Communicated by the Rev. Mr. Richard Price, in a letter to Mr. John Canton, M.A. F.R.S.," *Philosophical Transactions*, 54, 296-325.
- Bernstein. S. N. (1917). "On the axiomatic foundation of the theory of probability," (Russian) , *Communications of the Kharkov Mathematical Society*, 15, 209-274.
- Bernstein. S. N. (1927). *The Theory of Probabilities*, State Publishing House. Moscow, Leningrad. (Russian).
- Bernstein. S. N. (1934). *The Theory of Probabilities* (2nd edition, augmented) , State

- Technical and Theoretical Publishing House.Moscow, Leningrad. [The 4th edition, augmented (1946) .] (Russian).
- Bernstein. S. N. (1964). *Collected Works. Vol. IV: The Theory of Probability. Mathematical Statistics. [1911-1946]*, Moscow. (Russian).
- Bienaymé. I. J. (1838). “Mémoire sur la probabilité des résultats moyens des observations; démonstration directe de la règle de Laplace,” *Mémoires présentés par divers savants à l'Académie royale des sciences de l'Institut de France*, Tome 5, 513-558.
- Castillo. I. and R. Nickl. (2013). “Nonparametric Bernstein-von Mises theorems in Gaussian white noise,” *The Annals of Statistics*, 41 (4) 1999 - 2028.
- Chernozhukov. V and H. Hong. (2003). “An MCMC approach to classical estimation,” *Journal of Econometrics*, 115(2) , 293 - 346.
- Cramér. H. (1953). “Richard Von Mises' Work in Probability and Statistics,” *The Annals of Mathematical Statistics*, 24 (4) , 657 - 662.
- Czuber. E. (1899). “Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen,” *Jahresbericht der Deutschen Mathematiker-Vereinigung*, Bd. 7, Nr.2, 1-279.
- Dale. A. I. (2012). *A History of Inverse Probability: From Thomas Bayes to Karl Pearson*, Second Edition, Springer-Verlag New York.
- Diaconis. P. and D. Freedman. (1986). “On the Consistency of Bayes Estimates,” *The Annals of Statistics*, 14 (1) , 1 - 26.
- Edgeworth. F.Y. (1908). “On the Probable Errors of Frequency-Constants,” *Journal of the Royal Statistical Society*, 71(2) , 381 -397, 499 - 512, 651 - 668.
- Edwards, A.W.F. (1978). “Commentary on the Arguments of Thomas Bayes,” *Scandinavian Journal of Statistics*, 5(2) , 116-118.
- Fischer. H. (2011). *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*, Springer-Verlag New York.
- Freedman. D. (1999). “On the Bernstein-Von Mises Theorem with Infinite-Dimensional Parameters,” *The Annals of Statistics*, 27(4) , 1119-1140.
- Gauss. C. F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*, English translation: Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections, by C.H. Davis, 1857, Little, Brown and Company, Boston. Reprinted 1963, Dover, New York.
- Ghosal. S. and A. W. van der Vaart. (2017). *Fundamentals of Nonparametric Bayesian Inference*, Cambridge University Press.
- Ghosh. J. K. and R. V. Ramamoorthi. (2003). *Bayesian Nonparametrics*, Springer-

Verlag New York.

- Gorroochurn. P. (2016). *Classic Topics on the History of Modern Mathematical Statistics: From Laplace to More Recent Times*, John Wiley and Sons.
- Hald. A. (2010). *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935*, Springer-Verlag New York.
- Heyde. C. C. and E. Seneta. (1977). *I. J. Bienaymé: Statistical Theory Anticipated*, Springer-Verlag New York.
- Heyde. C. C. and E. Seneta eds. (2001). *Statisticians of the Centuries*, Springer-Verlag New York.
- Keynes. J. M. (1921). *A Treatise on Probability*, Macmillan, London. . Reprinted as Vol. VIII of The Collected Writings of John Maynard Keynes (1973).
- Kitagawa. T, J. L. Montiel Olea, J. Payne, A. Velez. (2020). “Posterior distribution of nondifferentiable functions,” *Journal of Econometrics*, 217(1) , 161-175.
- Kleijn. B. J. K and A. W. van der Vaart. (2012). “The Bernstein-Von Mises theorem under misspecification,” *Electronic Journal of Statistics*, 6, 354-381.
- Laplace. P. S. (1774). “Mémoire sur la probabilité des causes par les évènements,” *Mémoires de mathématique et de physique, présentés à l'Académie royale des sciences, par divers Savans, & lûs dans ses assemblées*, Tome 6, 621-656. Oeuvres complètes de Laplace, Tome 8, 27-65. (1891) Translated into English with an introduction by S.M. Stigler in *Statistical Science*, 1, 359-378 (1986) .
- Laplace. P. S. (1785). “Mémoire sur les approximations des formules qui sont fonctions de très grand nombres,” *Mémoires de mathématique et de physique, tirés des registres de l'Académie Royale des Sciences*, Année 1782. 1-88. Oeuvres complètes de Laplace, Tome 10, 209-291. (1894).
- Laplace. P. S. (1786). “Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres. (Suite),” *Mémoires de mathématique et de physique, tirés des registres de l'Académie Royale des Sciences*, Année 1783. 423-467. Oeuvres complètes de Laplace, Tome 10, 295-338. (1894).
- Laplace. P. S. (1810a). “Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités,” *Mémoires de l'Académie des Sciences*, Année 1809, 353 - 415. Oeuvres complètes de Laplace, Tome 12, 301 - 345.
- Laplace. P. S. (1810b). “Supplément au Mémoire sur les approximations des formules qui sont fonctions de très grands nombres,” *Mémoires de l'Académie des Sciences*, Année 1809, 555-565. Oeuvres complètes de Laplace, Tome 12, 349 - 353.

- Laplace. P. S. (1812). *Théorie Analytique des Probabilités*, Courcier, Paris. 2nd ed. (1814) , 3rd ed. (1820) . Quatrième Supplement (1825) . Oeuvres complètes de Laplace, Tome 7. (1886) (邦訳：「ラプラス確率論-確率の解析的理論-」、伊藤清、樋口順四郎訳、1986、共立出版.)
- Le Cam. L. (1953). “On some asymptotic properties of maximum likelihood estimates and related Bayes estimates,” *University of California Publications in Statistics*, 1, No.11, 277-330.
- Le Cam. L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York.
- Le Cam. L and G. L. Yang. (1990). *Asymptotics in Statistics: Some Basic Concepts*, 2nd edition (2012) . Springer, New York.
- Lehmann. E. L. and G. Casella. (1998). *Theory of Point Estimation*, second edition, Springer-Verlag, New York.
- Maistrov. L. E. (1967). *Probability Theory - A historical Sketch*, (Russian) Izdatel'stvo Nauka, Moscow. Translated into English and edited by S. Kotz. (1974) , Academic Press.
- Norets. A. (2015). “Bayesian regression with nonparametric heteroskedasticity,” *Journal of Econometrics*, 185(2) , 409-419.
- Pearson. K. and L.N.G. Filon. (1898). “Mathematical contributions to the theory of evolution, IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation,” *Philosophical Transactions*, Series A, 191, 229 - 311.
- Poisson. S. D. (1830). “Mémoire sur la proportion des naissances des filles et des garçons,” *Mémoires de l'Académie des sciences de l'Institut de France*, Tome IX, 239-308.
- Poisson. S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile ; précédées des Règles générales du calcul des probabilités*, Bachelier, Paris.
- Seneta. E. (2013). “A Tricentenary history of the Law of Large Numbers,” *Bernoulli*, 19 (4) , 1088-1121.
- Sheynin. O. (2017). *Theory of Probability. A Historical Essay* , revised and enlarged edition, unpublished paper.
- Stigler. S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*, Harvard University Press.
- Todhunter. I. (1865). *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, Cambridge University Press, London and

Cambridge.

- van der Vaart. A. W. (1998). *Asymptotic Statistics*, Cambridge University Press.
- Videnskii. V.S. (2000). “Academician Sergey Natanovich Bernstein (on the 120th anniversary of his birth) ,” *Bulletin of Young Scientists 4'2000*, Series: Applied Mathematics and Mechanics 2' 2000, 2-11.
- von Mises. R. (1919a). “Fundamentalsätze der Wahrscheinlichkeitsrechnung,” *Mathematische Zeitschrift*, 4, 1-97.
- von Mises. R. (1919b). “Grundlagen der Wahrscheinlichkeitsrechnung,” *Mathematische Zeitschrift*, 5, 52-99.
- von Mises. R. (1931). *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Vorlesungen aus dem Gebiete der Angewandten Mathematik, Bd. I, Leipzig u. Wien.
- von Mises. R. (1964). *Mathematical Theory of Probability and Statistics*, (Edited and complemented by H. Geiringer.) Academic Press, New York.