

水産資源研究のための数学特論

赤嶺 達郎*

Special Lecture of Mathematics for Fish Stock Analysis

Tatsuro AKAMINE*

Abstract : There are some mistakes for the negative binomial distribution and the log-binomial distribution in the economy and the ecology. The corrections of these are presented, and the Bayesian statistical methods using the highest density region and the probability matching prior distribution are explained. The mathematics that are the quadrat method, the hypergeometric distribution, the random walk model, the gamma and beta functions, the chi-squared distribution, are explained. These are backgrounds of statistical method of fisheries science, and knowledge of these helps us to study fish stock analysis.

キーワード : 負の二項分布, 対数正規分布, 最高確率密度区間, 確率一致事前分布, ランダム・ウォーク

水産資源研究における数理モデルおよび統計モデルについて、筆者は既にテキストを2冊出版しているが(赤嶺, 2007, 2010)、価格を押さえる必要から数学的に専門的な部分は割愛せざるを得なかった。そのような専門的ないくつかの項目や補足事項についてこの総説で解説する。それらについての解説書や専門書は数多く出版されており、インターネットでも容易に情報を入手できる時代である。しかし単に知識を得るだけでは使い物にならない。長沼(2011)にあるような直感的に理解できる明瞭な説明が必要である。また純粋数学では重要な事項であっても、応用上ほとんど必要ないものも多い。それらを見分ける能力を養う上からも、ある程度の知識は必要である。とりあえず先に進んで、全体を俯瞰するような態度も研究を遂行する上で重要である。

生態学的な知見を得るだけなら、数式がない渡辺(2012)のような良書を読むだけで十分である。しかし水産分野においてもデータを用いて研究する際には、統計的な検定や推定が不可欠となっている。マニュアル通りに解析すれば通用する分野もあるが、ほとんどの分野でそのデータごとに数理モデルを作成して厳密

な統計処理をすることが要求される。生態学の分野ではそのような教科書として島谷(2012)がある。ここでは数理モデルから統計モデルへ、AIC(赤池の情報量規準)からベイズ統計へという2つの流れが示されている。水産分野においても今後はこのような研究が主流になると期待される。

最初の第1章では負の二項分布と対数正規分布について解説する。確率についての初歩的な内容であるが、解釈を間違った書物が散見されるからである。次の第2章では赤嶺(2010)の補足として具体的な計算事例をいくつか解説する。水産分野では必要以上に数式アレルギーの人が多く見受けられる。この総説の後半では非常に多くの数式が出てくるが、眺めるだけでもよい。使う必要が出てくれば、自然に理解できるようになるからである。また他分野の研究者と共同研究を行う際にも、このような知識は有益となるだろう。

最近、生態学において確率微分方程式に基づく絶滅確率の推定方法が提示されており(Hakoyama and Iwasa, 2000)、水産資源についても検討する必要性が生じてきている。しかし確率微分方程式はルベーグ積分に関係していて、正確にイメージすることは難しい。

2013年9月5日受理 (Received on September 5, 2013)

*独立行政法人 水産総合研究センター 中央水産研究所 〒236-8648 神奈川県横浜市金沢区福浦2-12-4 (National Research Institute of Fisheries Science, Fisheries Research Agency, 2-12-4 Fukuura, Kanazawa, Yokohama 236-8648, Japan)

そこで確率や統計についての基礎的な概念や歴史的事項について付録で簡単に解説した。この分野の専門書は非常に難しく、逆に入門書は易しすぎる。長沼(2011)が指摘しているように、厳密さよりもイメージをつかむことの方が重要である。

第1章 負の二項分布と対数正規分布

二項分布はもっとも基本的な確率分布であり、これから派生する負の二項分布や対数正規分布も基本的な確率分布であるが、生態学関係において数学的な誤りが散見される。ここではこれらの誤りについて解説するとともに、関連してベイズ統計学におけるHDR (Highest Density Region: 最高確率密度区間) とPMP (Probability Matching Prior: 確率一致事前分布)、およびランダム・ウォークに関する基本事項について解説する。

負の二項分布

二項分布は成功率 p 、総試行数 n のとき、成功数 r の確率で、

$$\text{Bi}(r; n, p) = \binom{n}{r} p^r (1-p)^{n-r} \quad (1.1)$$

と定義される。ここでセミコロン (;) の左側が確率変数、右側がパラメータ (母数)、

$$\sum_{r=0}^n \text{Bi}(r; n, p) = 1 \quad (1.2)$$

である。なお

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n^{(r)}}{r!} \quad (1.3)$$

は組合せ数である。ここで $n^{(r)}$ は Aitken の記号と呼ばれ、

$$n^{(r)} = n(n-1) \cdots (n-r+1) \quad (1.4)$$

である (安藤, 2001)。

一方、負の二項分布は成功率 p のとき、 r 回成功するまでの総試行数 n の確率で、

$$\text{NB}(n; r, p) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad (1.5)$$

と定義され、

$$\sum_{n=r}^{\infty} \text{NB}(n; r, p) = 1 \quad (1.6)$$

である。これは二項分布と以下の関係がある。

$$\text{NB}(n+1; r+1, p) = \binom{n}{r} p^{r+1} (1-p)^{n-r} = p \text{Bi}(r; n, p) \quad (1.7)$$

これより次の総和公式が成立する (赤嶺, 1988)。

$$\sum_{i=0}^r \text{Bi}(i; n, p) = \sum_{j=r}^{\infty} \text{NB}(j+1; r+1, p) \quad (1.8)$$

これは竹内・藤野(1981)によると、ほとんど自明な公式で「 $r+1$ 回の成功が起こるまでに $n+1$ 回の試行を必要とするということは、 n 回目までの試行では高々 r 回の成功しか起こらなかったということ」を意味している。

ところで Sober (2008) は二項分布と負の二項分布における検定の違いを紹介している。 $n=20, r=6$ の場合に帰無仮説「 $H_0: p=0.5$ 」を検定すると、

(a) 二項分布では $r=0 \sim 6$ および $14 \sim 20$ の確率は 0.115。

(b) 負の二項分布では $n=20 \sim \infty$ となる確率は 0.0319。

これより $\alpha=0.05$ で (a) は棄却できないが、(b) は棄却できると判定される。これは奇妙な結果である。また確率の値が大きく異なるが、原因は何だろうか?

じつはこのような指摘は既に繁梈(1985)にある。二項分布における p の不偏推定量は r/n であるが (この場合は $6/20=0.3$)、負の二項分布における p の不偏推定量は $(r-1)/(n-1)$ である (この場合は $5/19=0.263$)。同じデータを得ても、調査手法およびモデルによって不偏推定が異なってしまうのである。前者はあらかじめ試行回数を 20 回に決めて試行したところ、たまたま 6 回成功した場合で、後者はあらかじめ成功数を 6 回と決めて試行したところ、たまたま 20 回目で終了した場合である。なお後者の不偏推定量は演習問題として Hoel (1971) や安藤・門脇 (2004) に提示されているので、以下に示す。

$$\begin{aligned} E\left(\frac{r-1}{n-1}\right) &= \sum_{n=r}^{\infty} \frac{r-1}{n-1} \text{NB}(n; r, p) \\ &= p \sum_{n=r}^{\infty} \binom{n-2}{r-2} p^{r-1} q^{n-r} = p \end{aligned} \quad (1.9)$$

この問題では成功率 p の検定を行うのに、前者では成功数 r の確率分布を利用し、後者では総試行数 n の確率分布を利用している。単純に考えれば、 p の推定値 r/n は二項分布では不偏であり、負の二項分布では偏っているから、この問題では二項分布を用いるのが妥当である。これは極端な例を考えると理解しやすい。 $p=0$ のときに二項分布では正しく推定されるが、負の二項分布では永遠に試行しなくてはなら

ない。また負の二項分布において $r = 1$ と設定して $n = 1$ というデータが得られたとき、これから $p = 1$ と推定するのはかなり乱暴である。

Table 1. Probability (A) and cumulative probability (B) of the binomial distribution when $n=20$ and $p=0.5$.

r	A	B
0	0.00000	0.00000
1	0.00002	0.00002
2	0.00018	0.00020
3	0.00109	0.00129
4	0.00462	0.00591
5	0.01476	0.02069
6	0.03696	0.05766
7	0.07393	0.13159

負の二項分布では「 n 回目は必ず成功」する。したがって p を検定する目的で負の二項分布を用いる場合には、「 n 回目のデータを除く」必要がある。表計算ソフトを用いて検算してみよう。二項分布では $n = 20$, $p = 0.5$ のとき **Table 1.** となる。 $r = 0 \sim 6$ の確率が 5.766% だから、片側 2.5% の棄却域では棄却できない。一方、負の二項分布では $r = 7$, $p = 0.5$ のとき **Table 2.** となる。

$n = 21 \sim \infty$ となる確率は $0.057659 (= 0.115318 / 2)$ となるから、(a) と結果が完全に一致する。ただし厳密に棄却域を考えるのであれば、 $n = 7$ の確率の方が $n = 21$ の確率よりも小さいので、棄却域は $n = 7, 21 \sim \infty$ として確率 0.0655 とすべきである。これは 5% よりも大きいから棄却できない。以上のように両者の検定結果は一致する。検定結果がほとんど一致するのは (1.8) 式が成立しているためである。これについてはベイズ統計の立場から「枠どり法」の節で検討する。

ベータ分布の利用

成功率 p を検定するためにベイズ統計を用いるのであれば、直接 p の確率分布を考えの方が簡単である。 p の事前分布を $p = 0 \sim 1$ における一様分布と仮定すると、事後分布はベータ分布となる。0.01きざみで表計算ソフトを用いて計算すると、**Table 3.** となる。こ

Table 2. Probability (A) and cumulative probability (B) of the negative binomial distribution when $r=7$ and $p=0.5$.

n	A	B	n	A	B
7	0.00781	0.00781	15	0.09164	0.69638
8	0.02734	0.03516	16	0.07637	0.77275
9	0.05469	0.08984	17	0.06110	0.83385
10	0.08203	0.17188	18	0.04721	0.88106
11	0.10254	0.27441	19	0.03541	0.91647
12	0.11279	0.38721	20	0.02588	0.94234
13	0.11279	0.50000	21	0.01848	0.96082
14	0.10474	0.60474	22	0.01294	0.97376

れより $p = 0.14 \sim 0.50$ の区間 (HDR) の確率が $96.09 - 2.02 = 94.07\%$ となる。したがってこの場合も $p = 0.50$ は 95% HDR 区間に含まれているので棄却できない。

p と r については

$$n \int_0^p \text{Bi}(r; n-1, t) dt = \sum_{i=r+1}^n \text{Bi}(i; n, p) \quad (1.10)$$

または

$$(n+1) \int_0^p \text{Bi}(r; n, t) dt = p \text{Bi}(r; n, p) + \sum_{i=r+1}^n \text{Bi}(i; n, p) \quad (1.11)$$

という関係がある。これは p の下側確率と r の上側確率がほぼ一致することを示している。Dey and Rao (2005) に「事後確率と被覆確率とが正確にもしくは近似的に一致するような事前確率」として PMP (Probability Matching Prior: 確率一致事前分布) が定義されている。したがって p の事前分布として一様分布を採用すれば PMP となる。

ベイズ統計では事後分布における区間推定に HPD (Highest Posterior Density Region: 最高事後密度領域) を採用することが多い (Dey and Rao, 2005)。これは繁樹 (1985) では最高密度信頼領域 (Highest

Credibility Density Region) と呼ばれ、渡部 (1999) では HDR (Highest Density Region: 最高密度領域) と呼ばれている。伝統的統計学では信頼区間の推定において便宜的に確率分布の片側 $a/2$ 点を用いることが多かったが、ベイズ統計における事後分布と同様に最高確率密度区間を採用する方がよい (区間の幅が最短となるから)。ここでは事後分布に限定しないで、一般の確率分布において HDR (Highest Density Region) と呼ぶことにする。このときベイズ統計の事後分布と伝統的統計学の信頼区間とで、片側確率が一致するという条件では両者の HDR は必ずしも一致しない (赤嶺, 2002)。したがって PMP は絶対的な条件ではなく、「望ましい性質の1つ」である (Dey and Rao, 2005)。

なお片側確率が一致する性質を用いて二項分布の「正確な検定」や推定を、 F 分布またはベータ分布で行う手法がある。同様にポアソン分布ではカイ2乗分布またはガンマ分布を用いる。証明は小寺 (1986) にある。片側検定や片側推定のときのみ正確で、HDR では一致しない。また松原 (2010) はベイズ統計の歴史に詳しいが、HDR や PMP について言及していない。

枠どり法

枠どり法は方形枠などを用いて採集した個体数から調査域全体の総個体数を推定する方法で、数学モデルとして部分観察法や抽出法と同一である。これと同じ考え方で、ある湖に生息する魚の総個体数 n を推定してみよう。漁獲率 p で漁獲したところ r 尾が漁獲されたとき、 n の確率分布を求める問題である。このとき n と p をパラメータ、 r を確率変数とするのが自然である。つまり枠どり法を何回も復元試行した場合、採集尾数 r は変動するが、総個体数 n は変動しない。したがって二項分布モデルを適用するのが妥当で、伝統的統計学では帰無仮説 $H_0: n = n_0$ を立てて信頼区間を推定する。一方、ベイズ統計では「 n の事前分布を一様分布と仮定」して n の事後分布を求める。これが負の二項分布と一致するわけである (赤嶺, 1988, 2002, 2007, 2010)。

このことは以下のように解釈できる。漁獲率 p と漁獲尾数 r は既知である。最初に湖の魚全部 (n 尾) を生け捕りして、1尾ずつ順番に確率 p で漁獲、確率 $1-p$ で再放流すると考える。このままでは最後の1尾を漁獲したか再放流したか不明なので負の二項分布を適応できない。そこで n 回試行した後、「 $n+1$ 回目に $r+1$ 尾目を確率 p で漁獲した」と考える。それより前は「 n 回の試行で r 尾が漁獲された」わけだから、 n の

Table 3. Probability (A) and cumulative probability (B) of the beta distribution when $r=6$ and $n=20$.

p	A	B
0.135	0.00647	0.020208
0.145	0.00844	0.028648
0.285	0.03981	0.40929
0.295	0.04020	0.44948
0.305	0.04020	0.48968
0.315	0.03982	0.52950
0.495	0.00840	0.96088
0.505	0.00716	0.96803

確率分布は負の二項分布 (1.7) 式で与えられる。これは事前分布「 n は $r \rightarrow \infty$ で一様分布」を仮定して通常の二項分布を用いるベイズ統計モデルの事後確率と同一である。

ここで n の事前分布を一様分布と仮定しているが、これは従来「無情報事前分布」と解釈されることが多かった。これは知識が欠けている状態を示す事前分布で、通常は一様分布が採用される (繁樺, 1985)。しかし赤嶺 (2007, 2010) で扱っているベイズ統計モデルは、この場合も含めすべて PMP (Probability Matching Prior) である。

超幾何分布

超幾何分布は

$$HG(r; n, M, N) = \frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}} \quad (1.12)$$

と定義され、

$$\sum_{r=0}^n HG(r; n, M, N) = 1 \quad (1.13)$$

である。Link and Barker (2010) には超幾何分布において M の事前分布を一様分布と仮定した場合の事後分布

$$\begin{aligned} \text{Post}(M; r, n, N) &= \frac{\frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}} \frac{1}{N+1}}{\sum_{M=0}^N \frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}} \frac{1}{N+1}} \\ &= \frac{\binom{M}{r} \binom{N-M}{n-r}}{\sum_{M=r}^{N-n+r} \binom{M}{r} \binom{N-M}{n-r}} \quad (1.14) \end{aligned}$$

が (6.4) 式として提示されていて、その下に数値表 (Table 6.1) も載っている。しかしこの式の分母は

$$\sum_{M=r}^{N-n+r} \binom{M}{r} \binom{N-M}{n-r} = \binom{N+1}{n+1} = \frac{N+1}{n+1} \binom{N}{n} \quad (1.15)$$

である (赤嶺, 2002)。したがって

$$\text{Post}(M; r, n, N) = \frac{n+1}{N+1} HG(r; n, M, N) \quad (1.16)$$

のように簡単に表せる。

一様分布が常に PMP (Probability Matching Prior) になるわけではない。超幾何分布の N については事前分布

$$\pi(N) = \frac{M+1}{(N+1)(N+2)} \quad (1.17)$$

が PMP である (赤嶺, 2002, 2007, 2010)。

対数正規分布

対数正規分布の定義は、確率変数 X の対数が正規分布に従うことで、

$$\ln X \sim N(\mu, \sigma^2) \quad (1.18)$$

と表わされる。 $y = \ln x$ とおくと、確率の変数変換公式 $P(y) dy = P(x) dx$ より、

$$P(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \quad (1.19)$$

および

$$\begin{aligned} P(x) &= P(y) \frac{dy}{dx} = \frac{P(\ln x)}{x} \\ &= \frac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] \quad (1.20) \end{aligned}$$

となる。これが対数正規分布の確率密度関数である。これより

$$\text{Mode}(x) = \exp(\mu - \sigma^2) \quad (1.21)$$

$$\text{Median}(x) = \exp(\mu) \quad (1.22)$$

$$E(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad (1.23)$$

$$V(x) = \exp(2\mu + \sigma^2) (\exp(\sigma^2) - 1) \quad (1.24)$$

を得る。

生態学で広く用いられている推移行列モデルでは、個体群サイズ N および最大固有値 λ は対数正規分布に従う。赤嶺 (2010) の第 7 章で推移行列モデルについて検討したが、これについては Akamine and Suda (2011) にまとめている。ここではより単純な 1 変数モデル

$$N(t+1) = N(t) \lambda(t) \quad (1.25)$$

で検討してみよう。両辺の対数をとると

$$\ln N(t+1) = \ln N(t) + r(t) \quad (1.26)$$

となる。ここで $r(t) = \ln \lambda(t)$ は内的増加率である。例として「平均値のパラドックス」をとりあげる (吉

村, 2012)。このモデルでは良い環境のとき $\lambda(t) = \alpha = 3.9$ で増加し、悪い環境のとき $\lambda(t) = \beta = 0.1$ で減少する。良い環境と悪い環境がランダムに起きると仮定すると、期待値は算術平均 $E = (\alpha + \beta)/2 = 2$ となるから増大するはずなのに、幾何平均が $\sqrt{\alpha\beta} = 0.6245$ となるため減少する、というパラドックスである。

じっさいに表計算ソフトを用いて $N(0) = 1$ のとき $N(5)$ を求めてみよう。これは

$$E^5 = \frac{(3.9 + 0.1)^5}{2^5} \quad (1.27)$$

を展開すればよい。結果を **Table 4.** に示す。一方、 $\ln N$ は「ランダム・ウォーク」なので二項分布に従うが、 $\ln \alpha = 1.361, \ln \beta = -2.303$ なので、

$$\mu = \frac{\ln \alpha + \ln \beta}{2} t, \quad \sigma^2 = \frac{(\ln \alpha - \ln \beta)^2}{4} t \quad (1.28)$$

の正規分布で近似できる。これより $t=5$ のとき、正規分布 $\text{Normal}(-2.354, 16.777)$ となる。じっさい **Table 4.** における $\ln N$ の標本分散は 16.777 である。この正規分布において変動係数は

$$CV = \frac{\sigma}{\mu} = \frac{1.8318\sqrt{t}}{0.4708t} = \frac{3.891}{\sqrt{t}} \quad (1.29)$$

となっている。したがって $t \rightarrow \infty$ のとき $CV \rightarrow 0$ となる。この正規分布について「連続補正」して表計算ソフトを用いて確率（面積）を求めると、**Table 5.** のよ

うになる。かなり良い近似である。表計算ソフトには累積正規分布が組込み関数として用意されているので、簡単に計算できる。

しかし対数正規分布では積分区間が大きく変化することに注意する必要がある。一見しただけでは対数正規分布が、対数値が二項分布に従う確率分布の近似になっているとは判断できない。ここで対数値が二項分布に従う確率分布を「対数二項分布」と呼ぶことにしよう。つまり

$$\ln X \sim \text{Bi}(n, p) \quad (1.30)$$

である。先ほどの公式を対数正規分布に用いると、

$$\begin{aligned} \text{平均} & e^{6.034} = 417.4 \\ \text{分散} & e^{12.07} (e^{16.78} - 1) = 3.363 \times 10^{12} \\ \text{メディアン} & e^{-2.354} = 0.09499 \\ \text{モード} & e^{-19.13} = 0.00000000492 \end{aligned}$$

を得る。一方、 N の度数分布 (**Table 4**) から計算すると、

$$\begin{aligned} \text{平均} & 32 \\ \text{標本分散} & 24499 \\ \text{メディアン} & (0.5932 + 0.01521)/2 = 0.3042 \\ \text{モード} & 0.5932, 0.01521 \end{aligned}$$

となっていて、ほとんど一致していない。これは対数二項分布と対数正規分布の差に原因がある。

従来は $\ln N$ が正規分布に従うため、正規分布で平均値を求めたり区間推定を行ったりしていた。しかしそ

Table 4. Values of the random walk model.

N	$\ln N$	frequency	probability
902.2	6.805	1	0.03125
23.13	3.141	5	0.15625
0.5932	-0.5222	10	0.3125
0.01521	-4.186	10	0.3125
0.00039	-7.849	5	0.15625
0.00001	-11.51	1	0.03125
total	1024	32	1
mean	32	-2.354	

Table 5. Values approximated to the normal distribution in the random walk model.

ln N	interval	N	interval	probability
6.805	4.973~8.637	902.2	144.5~5634	0.03317
3.141	1.310~4.973	23.13	3.704~144.5	0.14873
-0.5222	-2.354~1.310	0.5932	0.09499~3.704	0.31446
-4.1858	-6.018~-2.354	0.0152	0.002436~0.09499	0.31446
-7.849	-9.681~-6.018	0.00039	0.0000624~0.002436	0.14873
-11.513	-13.345~-9.681	0.00001	0.0000016~0.0000624	0.03317

うすると実際の対数二項分布を用いた場合と値が大きく異なってしまう場合がある。対数二項分布のグラフを表計算ソフトで作図して解析するほうがよい。なお幾何平均と算術平均の差の解釈は巖佐 (1990) に解説されているように「移動分散の適応的意味」と解釈するのが妥当である。成長率 λ の算術平均は対数正規分布の平均 (Mean), 幾何平均はメディアン

$$\sqrt[n]{\lambda_1 \cdots \lambda_n} = \exp \frac{\ln \lambda_1 + \cdots + \ln \lambda_n}{n} = e^\mu \quad (1.31)$$

である。両者の優劣を論じるよりも、全体の確率分布を把握することが重要である。また $(a, \beta) = (4.0), (3.732, 0.268), (3.7, 0.3),$ および $(2, 2)$ などの数値例を検討することによって理解を深めることができる。

第2章 具体的な計算事例

赤嶺 (2010) では分かりやすさを優先して対話形式としたため長くなりすぎ、いくつかの事項を割愛せざるを得なかった。それらは計算事例が中心で、専門的かつ別解法となるものであるが、これらを赤嶺 (2010) と比較検討することによって理解が深まると思われる。なお、この章で検討する確率分布はベイズ統計の普及とともに重要性が増してきているように思う。

線型代数の応用

昔の教科書では最初に行列式 (determinant) の定義

$$\sum \text{sgn } \sigma \cdot a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \quad (2.1)$$

が載っていて、ほとんど理解不能であった。最近では教

育の配慮がなされた教科書が出ている (梶原, 2010 など)。筆者は行列を高校で教わらなかった最後の卒業生だったが、最近また高校で教えなくなった (京極, 2011)。大学入試で難しい問題が出るためらしい。入試問題の解法ではケーリー・ハミルトンの定理がよく使われる。この定理の証明は線型代数の教科書を読んでも理解しにくいだが、笠原 (1982) に余因子行列を用いた簡明な証明が載っている。安藤 (2012) によると、高校では生徒よりも教師の方が行列嫌いらしい。

線型代数はイギリスのケーリーと、ドイツのグラスマンが独自に開発したが、当時はほとんど注目されなかった。1920年代後半に量子力学で使用され、大学での必須科目となった。固有 (eigen) 値という名称も物理学者であるディラックの命名である。統計学では多変量解析で使用されるが、ここでは例として主成分分析を解説する。

赤嶺 (2010) では主軸回帰の直線を座標変換で求めたが、通常は以下に示すように分散行列の固有値を用いる。分散行列 (を定数倍したもの) の固有値の定義は

$$\begin{pmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda_1 \begin{pmatrix} x \\ y \end{pmatrix} \quad (2.2)$$

である。これより第1行は

$$S_{xx}x + S_{xy}y = \lambda_1 x \quad (2.3)$$

となるから、主軸の傾きは

$$\frac{y}{x} = \frac{\lambda_1 - S_{xx}}{S_{xy}} \quad (2.4)$$

で与えられる。または第2行

$$S_{xy}x + S_{yy}y = \lambda_1 y \quad (2.5)$$

を解いて,

$$\frac{y}{x} = \frac{S_{xy}}{\lambda_1 - S_{yy}} \quad (2.6)$$

を得るが, これも同じ解を与える。(2.2) 式における固有方程式は,

$$\lambda^2 - (S_{xx} + S_{yy})\lambda + S_{xx}S_{yy} - S_{xy}^2 = 0 \quad (2.7)$$

となるから,

$$\lambda_1 = \frac{S_{xx} + S_{yy} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2} \quad (2.8)$$

を得る。これを (2.4) 式に代入すれば主軸の傾き b_1 を得る。この方法なら根号の前の符号を間違えない。しかし分散行列や固有値についての知識が必要である。なお主成分分析の統計パッケージでは固有値や固有ベクトルを数値計算で求めている。

補足事項

赤嶺 (2010) にはいくつかミスがある。まず「ヘッセ行列 (ヘシアン)」と書いたのは誤りではないが (p.26), ヘシアンは行列式 (determinant) に使うことが多いため, 「ヘッセ行列 (Hessian matrix)」のように修正したほうがよい。また関数方程式 (p.132)

$$f(x)f(y) = f(\sqrt{x^2 + y^2})f(0) \quad (2.9)$$

はガウスではなくて, 記号論理学で有名なブールのアイデアである (安藤, 1995)。なおベルヌイ方程式の解法はライプニッツによると書いたが (p.101), この出典は Thieme (2003) である。ヤコブ・ベルヌイが問題として提示し, ライプニッツとヨハン・ベルヌイが解答を与えたとのことである (森, 1973; Hairer and Wanner, 1996a)。

また平方根について

$$\sqrt{-a} = \sqrt{-1 \times a} = \sqrt{-1}\sqrt{a} = i\sqrt{a} \quad (2.10)$$

および

$$\sqrt{-a} = \sqrt{\frac{a}{-1}} = \frac{\sqrt{a}}{\sqrt{-1}} = \frac{\sqrt{a}}{i} = -i\sqrt{a} \quad (2.11)$$

という 2 通りの変形を記述した (p.141)。これは 2 次方程式 $x^2 + a = 0$ の解には 2 つの平方根 $x = \pm\sqrt{a}i$ が存在し, 区別が難しいためである。しかし根号には次のような規約がある。

- 1) $z > 0$ のとき, z の平方根の正の方を \sqrt{z} で表す。
- 2) $z < 0$ のとき, $z = -a$ とおくと, $\sqrt{-a} = \sqrt{a}i$ つまり $\sqrt{z} = \sqrt{-z}i$ と表す。

この規約に従うと $b < 0, c < 0$ のとき

$$\sqrt{bc} \neq \sqrt{b}\sqrt{c} \quad (2.12)$$

であり, $b < 0, c > 0$ のとき

$$\sqrt{\frac{c}{b}} \neq \frac{\sqrt{c}}{\sqrt{b}} \quad (2.13)$$

となる。したがって (2.10) 式が正しく, (2.11) 式は誤りである。

ここで複素数 $z = a + bi$ の平方根を上記の規約に合うように考える。

$$\sqrt{a + bi} = x + yi \quad (2.14)$$

とにおいて両辺を 2 乗すると,

$$a = x^2 - y^2, \quad b = 2xy \quad (2.15)$$

という連立方程式を得る。これを解いて, $b = 0$ のときに上記の規約に合うようにすると,

$$3) \quad b \geq 0 \text{ のとき, } \sqrt{z} = p + qi \text{ で表す。}$$

$$4) \quad b < 0 \text{ のとき, } \sqrt{z} = \pm(p - qi) \text{ で表す (+ と - のどちらでもよい)。}$$

となる。ここで

$$p = \sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}}, \quad q = \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}} \quad (2.16)$$

である (石谷, 1973)。この規約は複素平面上において極形式で考えると分かりやすい。 \sqrt{z} が第 1 象限にあるときだけ $b = 0$ としても根号の規約と合致し, それ以外の象限では上手くいかない。

複素数を水産資源学で用いることはほとんどない。しかし電磁気学や量子力学では不可欠であるし, フーリエ級数やラプラス変換は非常に重要な手法である。また複素関数を使うと, 素数定理を高校数学のレベルで証明できる (吉田, 2011)。なお素数定理は証明が難解なことである有名な定理で,

$$\lim_{x \rightarrow \infty} \frac{\pi(x) \ln x}{x} = 1$$

と表される (ここで $\pi(x)$ は x 以下の素数の個数)。

変数変換

中西ら (2003) に紹介されているリスク計算の公式を検討する。ダイオキシン類の体内濃度 (対数値) を x とする。このとき全人口における体内濃度の分布は正規分布, 感受性も正規分布で表せるので, リスク (確率) は,

$$P = \int_{-\infty}^{\infty} f(x)G(x)dx \quad (2.17)$$

と定義される。ここで

$$f(x) = N(\mu_1, \sigma_1^2)$$

$$G(x) = \int_{-\infty}^x g(y) dy$$

$$g(y) = N(\mu_2, \sigma_2^2)$$

である (N は正規分布, $\mu_1 < \mu_2$)。このとき公式

$$P = \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \quad (2.18)$$

が成立するが (Φ は累積標準正規分布), 以下にこの証明を考える。上式を具体的に書くと,

$$P = \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^x g(y) dy \right) dx \quad (2.19)$$

となる。これは二重積分で, 積分領域は $y \leq x$ である。つまり

$$P = \iint_D f(x)g(y) dx dy \quad (2.20)$$

である ($D: y \leq x$)。そこで変数変換

$$s = y - x \leq 0, \quad t = y + x \quad (2.21)$$

を考えると, 積分領域は「 $-\infty \leq s \leq 0, -\infty \leq t \leq \infty$ 」となるから簡単に積分できる。平均と分散の変換公式は

$$E(ax+by) = aE(x) + bE(y) \quad (2.22)$$

$$V(ax+by) = a^2V(x) + b^2V(y) \quad (2.23)$$

であるから,

$$s \approx N(\mu_2 - \mu_1, \sigma_1^2 + \sigma_2^2) \quad (2.24)$$

$$t \approx N(\mu_2 + \mu_1, \sigma_1^2 + \sigma_2^2) \quad (2.25)$$

となる。具体的には

$$\frac{s + \mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{y - \mu_2}{\sigma_2} - \frac{x - \mu_1}{\sigma_1} \quad (2.26)$$

$$\frac{t - \mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{y - \mu_2}{\sigma_2} + \frac{x - \mu_1}{\sigma_1} \quad (2.27)$$

と変数変換すればよい。そうすれば, $f(x)g(y) \approx \phi(s)\psi(t)$ を得る。ここで

$$\int_{-\infty}^{\infty} \psi(t) dt = 1 \quad (2.28)$$

だから, 残りの $\phi(s)$ を $s = -\infty \sim 0$ で積分すればよい。したがって

$$\int_{-\infty}^0 \phi(s) ds = \int_{-\infty}^m \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz,$$

$$m = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (2.29)$$

を得る。これが公式 (2.18) である。

このように一度に変換すると分かりにくいかもしれ

ないので, 別証明を考える。まず最初に

$$X = \frac{x - \mu_1}{\sigma_1}, \quad Y = \frac{y - \mu_2}{\sigma_2} \quad (2.30)$$

と変換すると, 直線 $y=x$ は直線

$$\sigma_2 Y + \mu_2 = \sigma_1 X + \mu_1 \quad (2.31)$$

に移る。また確率分布は2次元標準正規分布

$$P(X, Y) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(X^2 + Y^2)\right] \quad (2.32)$$

になる。これは原点に対して同心円となっているから, 先の直線と直交する直線 $Y=aX$ とその交点を求めればよい。ラグランジュの未定乗数法を適用すれば簡単である。直線 (2.31) 式上で $Z^2 = X^2 + Y^2$ の最小値を求めればよい。

$$Z^2 = X^2 + Y^2 + \lambda(\sigma_2 Y + \mu_2 - \sigma_1 X - \mu_1) \quad (2.33)$$

において,

$$\frac{\partial Z^2}{\partial X} = 2X - \lambda\sigma_1 = 0 \quad \text{より} \quad X = \frac{\lambda\sigma_1}{2} \quad (2.34)$$

$$\frac{\partial Z^2}{\partial Y} = 2Y + \lambda\sigma_2 = 0 \quad \text{より} \quad Y = -\frac{\lambda\sigma_2}{2} \quad (2.35)$$

となるから, (2.31) 式に代入して,

$$\lambda = \frac{2(\mu_2 - \mu_1)}{\sigma_1^2 + \sigma_2^2} \quad (2.36)$$

を得る。したがって

$$X = \frac{\sigma_1(\mu_2 - \mu_1)}{\sigma_1^2 + \sigma_2^2}, \quad Y = -\frac{\sigma_2(\mu_2 - \mu_1)}{\sigma_1^2 + \sigma_2^2} \quad (2.37)$$

となる。これより

$$Z^2 = X^2 + Y^2 = \frac{(\mu_2 - \mu_1)^2}{\sigma_1^2 + \sigma_2^2} \quad (2.38)$$

となるから,

$$Z = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} < 0 \quad (2.39)$$

を得る。これが求める交点の座標, つまり m の値である。

ガンマ関数

階乗を一般化したガンマ関数 $\Gamma(s)$ について多くの数学的に重要な事項が発生するのは, 驚くべきことである。ガンマ関数だけで1冊の本が書けるが, 逆にガンマ関数を素材として微分積分学の基礎を学ぶことができる (Artin, 1931)。ガンマ分布について知っておくことは無駄ではない。

ここでガンマ関数を歴史に沿って眺めてみる。オイラーは最初に x と n を正の整数として, 等式

$$(x-1)! = \frac{1 \cdot 2 \cdots (n-1)n^x}{x(x+1)(x+2) \cdots (x+n-1)} \times \left[\frac{n}{n} \frac{n+1}{n} \cdots \frac{n+x-1}{n} \right] \quad (2.40)$$

から

$$(x-1)! = \lim_{n \rightarrow \infty} \frac{(n-1)!n^x}{x(x+1)(x+2) \cdots (x+n-1)} \quad (2.41)$$

を導いた。右辺は x が整数でなくても成立する。この式からオイラーは

$$(x-1)! = \frac{1}{x} \prod_{n=1}^{\infty} \left(1 + \frac{1}{n}\right)^x \left(1 + \frac{x}{n}\right)^{-1} \quad (2.42)$$

という無限積を導いている（一松, 1972）。実際オイラーは1729年の手紙に

$$x! = \frac{1 \cdot 2^x}{1+x} \times \frac{2^{1-x} \cdot 3^x}{2+x} \times \frac{3^{1-x} \cdot 4^x}{3+x} \times \frac{4^{1-x} \cdot 5^x}{4+x} \times \cdots \quad (2.43)$$

と書き、この無限積より

$$\begin{aligned} \frac{1}{2}! &= \frac{1 \cdot \sqrt{2}}{3/2} \times \frac{\sqrt{2} \cdot \sqrt{3}}{5/2} \times \frac{\sqrt{3} \cdot \sqrt{4}}{7/2} \times \frac{\sqrt{4} \cdot \sqrt{5}}{9/2} \times \cdots \\ &= \sqrt{\frac{2 \cdot 4}{3 \cdot 3} \times \frac{4 \cdot 6}{5 \cdot 5} \times \frac{6 \cdot 8}{7 \cdot 7} \times \frac{8 \cdot 10}{9 \cdot 9} \times \cdots} \\ &= \sqrt{\frac{\pi}{4}} = \frac{\sqrt{\pi}}{2} \end{aligned} \quad (2.44)$$

を求めている（Dunhum, 2005）。ここでウォリスの公式

$$\frac{2 \cdot 2}{1 \cdot 3} \times \frac{4 \cdot 4}{3 \cdot 5} \times \frac{6 \cdot 6}{5 \cdot 7} \times \frac{8 \cdot 8}{7 \cdot 9} \times \cdots = \frac{\pi}{2} \quad (2.45)$$

または

$$\frac{2 \cdot 4}{3 \cdot 3} \times \frac{4 \cdot 6}{5 \cdot 5} \times \frac{6 \cdot 8}{7 \cdot 7} \times \cdots = \frac{\pi}{4} \quad (2.46)$$

を用いた。

結果に π が現れたので、オイラーは「何か丸い曲線の面積と関係している」と考え、1730年代の初頭に

$$n! = \int_0^1 (-\ln x)^n dx \quad (2.47)$$

を得た。これを置換すると、

$$n! = \int_0^{\infty} t^n e^{-t} dt \quad (2.48)$$

を得る。ルジャンドルがガンマ関数を

$$\Gamma(s) = \int_0^{\infty} t^{s-1} e^{-t} dt \quad (2.49)$$

と定義したが、この式はオイラー自身も導いている。

一方、ガウスはオイラーに従ってガンマ関数を

$$\Gamma(s) = \lim_{n \rightarrow \infty} \frac{n!n^s}{s(s+1) \cdots (s+n)} \quad (2.50)$$

と定義した。これはルジャンドルの定義式に指数関数の定義式

$$e^t = \lim_{n \rightarrow \infty} \left(1 + \frac{t}{n}\right)^n \quad (2.51)$$

を代入して、

$$\begin{aligned} \Gamma(s) &\approx \int_0^n t^{s-1} \left(1 - \frac{t}{n}\right)^n dt = n^s \int_0^1 x^{s-1} (1-x)^n dx \\ &= n^s B(s, n+1) = \frac{n^s \Gamma(s) \Gamma(n+1)}{\Gamma(n+s+1)} \end{aligned} \quad (2.52)$$

と変形しても得られる（笠原, 1973）。ガウスはガンマ関数を複素数にまで拡張している。さらに

$$\begin{aligned} \frac{1}{\Gamma(s)} &= \lim_{n \rightarrow \infty} \frac{s(s+1) \cdots (s+n)}{n!n^s} \\ &= \lim_{n \rightarrow \infty} s^{-s} \prod_{k=1}^n \left(1 + \frac{s}{k}\right) \end{aligned} \quad (2.53)$$

と変形すると、最後の無限積が収束しないので、

$$n^{-s} = e^{-s \ln n} \quad (2.54)$$

およびオイラー定数

$$\ln n \approx 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \gamma \quad (2.55)$$

を使って、

$$\frac{1}{\Gamma(s)} = s e^{\gamma s} \prod_{n=1}^{\infty} \left(1 + \frac{s}{n}\right) e^{-s/n} \quad (2.56)$$

とすれば収束する。これが有名なワイエルシュトラスの公式である（一松, 1963; Havil, 2003; 中田, 2005）。

どうしてこのような複雑な定義式が提唱されているのかというと、ガンマ関数には重要な公式がいくつかあり、これらの定義式を使うとそれらが自然に証明できるからで、より自然な定義式とみなせるのである（一松, 1990）。

カイ二乗分布

物理学者のファインマンは常に別証明を求めた（吉田, 2000）。赤嶺（2010）の第11章では小針（1973）に従ってカイ二乗分布と F 分布を導出したが、どうしてガンマ関数やベータ関数が出てくるのか分かりにくかった。小針（1973）では「数学的帰納法」と「たたみ込み」を用いているが、他にラプラス変換（積率母関数）を用いる方法もある（梶原, 1988）。ここでは宇喜多（1982）が2次元で解説した方法に従って、 n 次元の球（超球）の表面積を用いる方法を解説する。志村（2010）は「4次元の球体」

$$x_1^2+x_2^2+x_3^2+x_4^2 \leq r^2 \tag{2.57}$$

の体積の公式を学生に予想させるよう書いている。大学1年生向けの重積分の問題である。まず2次元の円の面積を求めてみる。「高校数学」のやり方でやってみると、

$$V(2) = S = 2 \int_0^r 2y dx = 4 \int_0^r \sqrt{r^2 - x^2} dx \tag{2.58}$$

ここで「 $x=r\sin\theta$ 」とおくと、「 $dx=r\cos\theta d\theta$ 」だから、

$$S = 4r^2 \int_0^{\pi/2} \cos^2\theta d\theta = 4r^2 \frac{1}{2} \frac{\pi}{2} = \pi r^2 \tag{2.59}$$

となる。ここで「ウォリス積分」の公式を用いた。同様に球の体積は、

$$\begin{aligned} V(3) &= 2 \int_0^r \pi y^2 dx = 2\pi \int_0^r (r^2 - x^2) dx \\ &= 2\pi r^3 \int_0^{\pi/2} \cos^3\theta d\theta = \frac{4}{3} \pi r^3 \end{aligned} \tag{2.60}$$

となる。この計算を繰り返すと、

$$\begin{aligned} V(4) &= 2 \int_0^r \frac{4}{3} \pi y^3 dx = \frac{8\pi}{3} \int_0^r (r^2 - x^2)^{3/2} dx \\ &= \frac{8\pi r^4}{3} \int_0^{\pi/2} \cos^4\theta d\theta = \frac{1}{2} \pi^2 r^4 \end{aligned} \tag{2.61}$$

を得る。なお1次元で同様の計算をすると、

$$V(1) = 2 \int_0^r (r^2 - x^2)^0 dx = 2 \int_0^r 1 dx = 2r \tag{2.62}$$

となる。これは直径の長さである。

以上より、漸化式

$$V(n) = 2rw(n) V(n-1) \tag{2.63}$$

を得る。ここで

$$w(n) = \int_0^{\pi/2} \cos^n\theta d\theta \tag{2.64}$$

はウォリス積分である。これより

$$V(n) = 2^n r^n w(n) w(n-1) \cdots w(1) V(0), \quad V(0) = 1 \tag{2.65}$$

となる。ウォリス積分は、

$$\begin{aligned} 2w(k) &= B\left(\frac{k+1}{2}, \frac{1}{2}\right) = \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{k}{2}+1\right)} \\ &= \sqrt{\pi} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}+1\right)} \end{aligned} \tag{2.66}$$

と表せるので (赤嶺, 2010),

$$\begin{aligned} V(n) &= (\sqrt{\pi})^n \frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)\Gamma\left(\frac{n+1}{2}\right)} \cdots \frac{\Gamma(1)}{\Gamma\left(\frac{3}{2}\right)} r^n \\ &= \frac{(\sqrt{\pi})^n}{\Gamma\left(\frac{n}{2}+1\right)} r^n = \frac{2(\sqrt{\pi})^n}{n\Gamma\left(\frac{n}{2}\right)} r^n \end{aligned} \tag{2.67}$$

を得る。これが求めていた n 次元の球の体積である。ここで両辺を微分すると

$$\Delta V = \frac{2(\sqrt{\pi})^n}{\Gamma\left(\frac{n}{2}\right)} r^{n-1} \Delta r \tag{2.68}$$

を得る。これが球の表面積である。

さて自由度 n のカイ二乗分布の定義は、

$$Y = X_1^2 + X_2^2 + \cdots + X_n^2 \tag{2.69}$$

である。このままでは分かりにくい、

$$y = r^2 = x_1^2 + x_2^2 + \cdots + x_n^2 \tag{2.70}$$

とおくと、これは半径 r の球の方程式である。したがって球の表面積を用いれば、カイ二乗分布が簡単に導ける。ここで確率変数の変換式を

$$P(y) \Delta y = P(x_1, x_2, \dots, x_n) \Delta V \tag{2.71}$$

と表わしてみる。標準正規分布の確率密度は

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{2.72}$$

なので、

$$\begin{aligned} P(x_1, x_2, \dots, x_n) &= P(x_1)P(x_2) \cdots P(x_n) \\ &= \frac{1}{(\sqrt{2\pi})^n} e^{-y/2} \end{aligned} \tag{2.73}$$

となる。

一方、 $y=r^2$ より、 $dy=2rdr$ なので、

$$\frac{\Delta V}{\Delta y} = \frac{\Delta V}{\Delta r} \frac{\Delta r}{\Delta y} = \frac{2(\sqrt{\pi})^n}{\Gamma\left(\frac{n}{2}\right)} r^{n-1} \frac{1}{2r} \tag{2.74}$$

となる。 $n=1$ のとき $\Delta V/\Delta r=2$ であるが、これは2価関数に対応している。以上より、

$$\begin{aligned}
 P(y) &= P \frac{\Delta V}{\Delta y} = \frac{e^{-y/2} (\sqrt{\pi})^n}{(\sqrt{2\pi})^n \Gamma\left(\frac{n}{2}\right)} r^{n-2} \\
 &= \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} y^{n/2-1} e^{-y/2} \quad (2.75)
 \end{aligned}$$

を得るが、これがカイ二乗分布の確率密度関数である。

ガンマ分布とベータ分布

次に分布の導出について考える。ここではガンマ分布とベータ分布に関する有名な定理を用いる。ガンマ関数の「ラプラス変換」型の定義は

$$\Gamma(s) = \int_0^{\infty} t^{s-1} e^{-t} dt \quad (2.76)$$

である。これを置換積分すると、

$$\int_0^{\infty} t^{s-1} e^{-\alpha t} dt = \frac{\Gamma(s)}{\alpha^s} \quad (2.77)$$

となる。これよりガンマ分布を

$$Ga(\alpha, \beta) = \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x} = \frac{\alpha^\beta}{\Gamma(\beta)} g(x, \beta) \quad (2.78)$$

と定義する(本によってパラメータの表現が異なる)。

これよりカイ二乗分布は

$$\chi_n^2(x) = Ga\left(\frac{1}{2}, \frac{n}{2}\right) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2} \quad (2.79)$$

と表せる。またベータ分布を

$$\begin{aligned}
 Be(p, q) &= \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1} \\
 &= \frac{1}{B(p, q)} b(x, p, q) \quad (2.80)
 \end{aligned}$$

と定義する。このとき次の定理が成立する。

[定理] X と Y が独立で、それぞれ $Ga(a, p)$ と $Ga(a, q)$ に従うとき、

(a) $X+Y$ は $Ga(a, p+q)$ に従う。

(b) $\frac{Y}{X+Y}$ は $Be(q, p)$ に従う。

この (b) から F 分布が導ける。Hairer and Wanner (1996b) を参考にすると、

$$\begin{aligned}
 I &= g(x, p) g(y, q) = x^{p-1} e^{-ax} y^{q-1} e^{-ay} \\
 &= x^{p-1} y^{q-1} e^{-a(x+y)} \quad (2.81)
 \end{aligned}$$

となるから、最初に変数変換

$$s = x + y, \quad v = y \quad (2.82)$$

を用いてみる。これは「正方形を同じ面積の平行四辺形に写す」写像なので、ヤコビアンは1である。した

がって

$$I = (s-v)^{p-1} v^{q-1} e^{-as} \quad (2.83)$$

となる。ここでさらに

$$v = st \quad (2.84)$$

と変換する。 $dv = s dt$ に注意すると、

$$\begin{aligned}
 I &= s^{p-1} (1-t)^{p-1} s^{q-1} t^{q-1} e^{-as} s \\
 &= s^{p+q-1} e^{-as} (1-t)^{p-1} t^{q-1} \\
 &= g(s, p+q) b(t, q, p) \quad (2.85)
 \end{aligned}$$

を得る。結局、分かりやすい変数変換は

$$s = x + y, \quad t = \frac{y}{x + y} \quad (2.86)$$

である。これは

$$y = -x + s, \quad \frac{x}{y} = \frac{1}{t} - 1 \quad (2.87)$$

という2種類の直線群である。以上より、

$$x = s - st = s(1-t), \quad y = st \quad (2.88)$$

となるから、ヤコビアンは

$$J = \det \begin{pmatrix} 1-t & -s \\ t & s \end{pmatrix} = (1-t)s + ts = s > 0 \quad (2.89)$$

である。これからガンマ関数とベータ関数の関係式

$$\Gamma(p) \Gamma(q) = \Gamma(p+q) B(p, q) \quad (2.90)$$

が簡単に証明できる。ガンマ分布とベータ分布に関する先の定理もまったく同様である。

一方、

$$T = \frac{Y}{X+Y} = \frac{1}{\frac{X}{Y} + 1} = \frac{1}{\frac{m}{n} U + 1} \quad (2.91)$$

とおくと、

$$U = \frac{X/m}{Y/n} = \frac{n}{m} \frac{X}{Y} = \frac{n}{m} \left(\frac{1}{T} - 1 \right) \quad (2.92)$$

となる。これより

$$dU = -\frac{n}{m} \frac{1}{T^2} dT \quad (2.93)$$

となるが、符号がマイナスなのは積分の方向が逆になるためである。

$$\frac{\Delta T}{\Delta U} = \left| \frac{dT}{dU} \right| = \frac{m}{n} T^2 \quad (2.94)$$

とすれば、 F 分布の確率密度関数

$$\begin{aligned}
 f(u) &= P \frac{\Delta T}{\Delta U} = \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \left(\frac{1}{1 + \frac{m}{n}u} \right)^{n/2-1} \\
 &\quad \times \left(1 - \frac{1}{1 + \frac{m}{n}u} \right)^{m/2-1} \frac{\frac{m}{n}}{\left(1 + \frac{m}{n}u\right)^2} \\
 &= \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \left(\frac{m}{n}\right)^{m/2} \frac{u^{m/2-1}}{\left(1 + \frac{m}{n}u\right)^{(m+n)/2}} \quad (2.95)
 \end{aligned}$$

が得られる。

赤嶺 (2010) の p.164 に書いたように, t 分布の t^2 は $F(1, n)$ に従うから, t 分布の確率密度を $p(t)$, $F(1, n)$ の確率密度を $f(u)$ とおくと, $t^2 = u$ だから, 2 価関数であることを注意して,

$$p(t) = f(u) t = f(t^2) t \quad (2.96)$$

を用いると,

$$\begin{aligned}
 p(t) &= \frac{1}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(\frac{1}{n}\right)^{1/2} \frac{t^{-1}t}{\left(1 + \frac{1}{n}t^2\right)^{(1+n)/2}} \\
 &= \frac{1}{\sqrt{n}B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \quad (2.97)
 \end{aligned}$$

を得る。

なお F 分布を求めるだけなら,

$$s = \frac{x}{y}, \quad t = x + y \quad (2.98)$$

とおくのがよい。これより

$$x = \frac{st}{1+s}, \quad y = \frac{t}{1+s} \quad (2.99)$$

となるから,

$$J = \det \begin{pmatrix} \frac{t}{(1+s)^2} & \frac{s}{1+s} \\ -\frac{t}{(1+s)^2} & \frac{1}{1+s} \end{pmatrix} = \frac{t}{(1+s)^2} > 0 \quad (2.100)$$

である (一松, 2011)。これを用いると,

$$\begin{aligned}
 I &= g(x, p)g(y, q) = x^{p-1}e^{-\alpha x}y^{q-1}e^{-\alpha y} \\
 &= t^{p+q-1}e^{-\alpha t} \frac{s^{q-1}}{(1+s)^{p+q}} \\
 &= g(t, p+q)h(s, q, p) \quad (2.101)
 \end{aligned}$$

を得る。ここで

$$v = \frac{s}{1+s} \quad (2.102)$$

とおくと,

$$1-v = \frac{1}{1+s}, \quad dv = \frac{ds}{(1+s)^2} \quad (2.103)$$

なので,

$$b(v) = h(s) \frac{ds}{dv} = \frac{s^{q-1}}{(1+s)^{p+q-2}} = v^{q-1}(1-v)^{p-1} \quad (2.104)$$

を得る。 $h(s)$ から F 分布の密度関数 $f(u)$ を求めるには,

$$u = \frac{n}{m}s \quad (2.105)$$

とおけばよい。

ま と め

応用数学に限らず, 何が重要で何が重要でないかを見極めることは重要であり, そのための眼力を養う必要がある。負の二項分布や対数正規分布のような初歩的な事項をしっかりと理解することが肝心である。統計学の基礎は重要であるが, 伝統的統計学における帰無仮説を用いた検定方法, およびベイズ統計学に関係した HDR (Highest Density Region: 最高確率密度区間) と PMP (Probability Matching Prior: 確率一致事前分布) を理解すれば, 後は自然に身につくように思う。

謝 辞

本論をまとめるにあたり協力していただいた (故) 須田真木さんと山越友子さん, また構成等について有益なコメントをいただいた査読者の方々に深謝いたします。

文 献

- 赤嶺達郎, 1988: 抽出法による個体数推定の誤差 (後編). 日水研連絡ニュース (345), 6-11.
- 赤嶺達郎, 2002: 枠どり法と Petersen 法の区間推定における伝統的統計学とベイズ統計学との比較. 水研センター研究報告, 2, 25-34.
- 赤嶺達郎, 2007: 水産資源解析の基礎. 恒星社厚生閣, 東京, 115pp.
- 赤嶺達郎, 2010: 水産資源のデータ解析入門. 恒星社厚生閣, 東京, 178pp.
- Akamine T. and Suda M., 2011: The growth rates

- of population projection matrix models in random environments. *Aqua-Bio Science Monographs*, 4, 95-104.
- 安藤洋美, 1989: 統計学けんか物語. 海鳴社, 東京, 142pp.
- 安藤洋美, 1995: 最小二乗法の歴史. 現代数学社, 京都, 240pp.
- 安藤洋美, 1997: 多変量解析の歴史. 現代数学社, 京都, 208pp.
- 安藤洋美, 2001: 大道を行く高校数学 統計数学編. 現代数学社, 京都, 429pp.
- 安藤洋美, 2012: 異説数学教育史. 現代数学社, 京都, 227pp.
- 安藤洋美, 門脇光也, 1980: ピアソンとフィッシャーの喧嘩物語 (II). BASIC 数学 (現代数学社), 1980年12月号, 32-35.
- 安藤洋美, 門脇光也, 1981: ピアソンとフィッシャーの喧嘩物語 (IV). BASIC 数学 (現代数学社), 1981年3月号, 36-40.
- 安藤洋美, 門脇光也, 2004: 初学者のための統計教室. 現代数学社, 京都, 194pp.
- Artin E., 1931: ガンマ関数入門 (上野健爾訳). 日本評論社, 東京, 126pp.
- Dey D.K. and Rao C.R., 2005: ベイズ統計分析ハンドブック (繁榊算男ら訳). 朝倉書店, 東京, pp.93-118.
- Dunham W., 2005: 微積分名作ギャラリー (一楽重雄ら訳). 日本評論社, 東京, 237pp.
- Geiringer H., 1968: 確率・客観的確率論 (安藤洋美訳). 歴史の中の数学, 平凡社, 東京, pp.114-197.
- Gessen M., 2009: 完全なる証明 100万ドルを拒否した数学者 (青木 薫訳). 文藝春秋, 東京, 375pp.
- Hairer E. and Wanner G., 1996a: 解析教程 上 (かに江幸博訳). シュプリンガー・フェアラーク東京, 東京, 323pp.
- Hairer E. and Wanner G., 1996b: 解析教程 下 (かに江幸博訳) シュプリンガー・フェアラーク東京, 東京, 339pp.
- Hakoyama H. and Iwasa Y., 2000: Extinction risk of a density-dependent population estimated from a time series of population size. *J. Theor. Biol.*, 204, 337-359.
- Havil J., 2003: オイラーの定数ガンマ γ で旅する数学の世界 (新妻弘ら訳). 共立出版, 東京, 284pp.
- 一松 信, 1963: 解析学序説 下. 裳華房, 東京, 316pp.
- 一松 信, 1972: ガンマ関数. 数学100の発見, 数学セミナー増刊, 日本評論社, pp.84-85.
- 一松 信, 1990: 微分積分学入門第三課. 近代科学社, 東京, 183pp.
- 一松 信, 2011: 多変数の微分積分学. 現代数学社, 京都, 271pp.
- Hoel P.G., 1971: 入門数理統計学 (浅井 晃, 村上正康訳). 培風館, 東京, 404pp.
- 石谷 茂, 1973: ルートの中は虚数でもよいか. \forall と \exists に泣く, 現代数学社, 京都, pp.124-129.
- 伊藤 清, 2004: 確率論の基礎 (新版). 岩波書店, 東京, 142pp.
- 伊藤清三, 1963: ルベーク積分入門. 裳華房, 東京, 301pp.
- 巖佐 庸, 1990: 数理生物学入門 生物社会のダイナミックス. HBJ 出版局, 東京, pp.192-205.
- 梶原じょう二, 1988: 新修文系・生物系の数学. 現代数学社, 京都, 268pp.
- 梶原 健, 2010: 基礎からわかる! しっかりわかる!! 線形代数ゼミ. ナツメ社, 東京, 320pp.
- 笠原こう司, 1973: 対話・微分積分学. 現代数学社, 京都, 344pp.
- 笠原こう司, 1982: 微分方程式の基礎. 朝倉書店, 東京, 207pp.
- 小針あき宏, 1973: 確率・統計入門. 岩波書店, 東京, 300pp.
- 小寺平治, 1986: 明解演習 数理統計. 共立出版, 東京, 212pp.
- Kolmogoroff A., 1974: 確率論の基礎概念 (第二版) (根本伸司訳). 東京図書, 東京, 118pp.
- 好田順治, 1984a: 微積分学史 (13) 測度論の展開. BASIC 数学 (現代数学社), 1984年2月号, 93-97.
- 好田順治, 1984b: 微積分学史 (14) ルベーク積分論からの発展. BASIC 数学 (現代数学社), 1984年3月号, 88-93.
- 京極一樹, 2011: ちょっとわかればこんなに役に立つ 中学・高校数学のほんとうの使い道. 実業之出版社, 東京, 207pp.
- Laplace P.S., 1812: 確率論 (伊藤清, 樋口順四郎訳). 共立出版, 東京, 442pp.
- Laplace P.S., 1814: 確率の哲学的試論 (内井惣七訳). 岩波書店, 東京, 287pp.
- Link W. A. and Barker R. J., 2010: Bayesian Inference with Ecological Applications. Elsevier, London, pp.111-113.
- 松原 望, 2010: ベイズ統計学概説 フィッシャーからベイズへ. 培風館, 東京, 255pp.

- 蓑谷千風彦, 1985: 回帰分析のはなし. 東京図書, 東京, 325pp.
- 蓑谷千風彦, 1987: 統計学のはなし. 東京図書, 東京, 293pp.
- 蓑谷千風彦, 1988: 推定と検定のはなし. 東京図書, 東京, 292pp.
- 森 真, 2012: 入門 確率解析とルベグ積分. 東京図書, 東京, 235pp.
- 森 毅, 1973: 異説数学者列伝. 蒼樹書房, 東京, pp.83-89.
- 長沼伸一郎, 2011: 物理数学の直感的方法 (普及版) 理工系で学ぶ数学の「難所突破」の特効薬. 講談社, 東京, 300pp.
- 中西準子, 益永茂樹, 松田裕之, 2003: 演習 環境リスクを計算する. 岩波書店, 東京, 230pp.
- 中田英樹, 2005: 社会人と大学生のための高校数学精義. 現代数学社, 京都, 246pp.
- 日本生態学会, 2004: 生態学入門. 東京化学同人, 東京, 273pp.
- 日本水産学会水産教育推進委員会, 2011: 農学・水産学系学生のための数理科学入門. 恒星社厚生閣, 東京, 137pp.
- 西村欣也, 2012: 生態学のための数理的方法 考えながら学ぶ個体群生態学. 文一総合出版, 東京, 261pp.
- 西山陽一, 2011: マルチンゲール理論による統計解析. 近代科学社, 東京, 168pp.
- Reid C., 1982: 数理統計学者ネイマンの生涯 (安藤洋美ら訳). 現代数学社, 京都, 525pp.
- 志賀浩二, 1990: ルベグ積分30講. 朝倉書店, 東京, 243pp.
- 繁榊算男, 1985: ベイズ統計入門. 東京大学出版会, 東京, 225pp.
- 島谷健一郎, 2012: フィールドデータによる統計モデリングとAIC. 近代科学社, 東京, 216pp.
- 志村五郎, 2010: 数学をいかに使うか. 筑摩書房, 東京, 173pp.
- Sober E., 2008: 科学と証拠 統計の哲学入門 (松王政浩訳). 名古屋大学出版会, 名古屋, 244pp.
- 高木貞治, 2010: 数学の自由性. 筑摩書房, 東京, 346pp.
- 竹内 啓, 藤野和建, 1981: 2項分布とポアソン分布. 東京大学出版会, 東京, 262pp.
- 竹内 啓, 大橋靖雄, 1981: 統計的推測 2標本問題. 数学セミナー増刊, 日本評論社, 東京, 192pp.
- Thieme H.R., 2003: 生物集団の数学 上 (齋藤保久訳). 日本評論社, 東京, 286pp.
- Todhunter L., 1865: 確率論史 (安藤洋美訳). 現代数学社, 京都, 531pp.
- 宇喜多義昌, 1982: 確率関数とそのグラフの応用. BASIC 数学 (現代数学社), 1982年6月号, 58-65.
- 渡部 洋, 1999: ベイズ統計学入門. 福村出版, 東京, 249pp.
- 渡辺 守, 2012: 生態学のレッスン 身近な言葉から学ぶ. 東京大学出版会, 東京, 185pp.
- 吉田耕作, 1977: 或る数学者の弁明 完備性問答. 数の直感にはじまる (彌永昌吉監修), 工作舎, 東京, 133-144.
- 吉田信夫, 2011: 複素解析の神秘性 複素数で素数定理を証明しよう!. 現代数学社, 京都, 205pp.
- 吉田 武, 2000: 虚数の情緒 中学生からの全方位独学法. 東海大学出版会, 神奈川, 1001pp.
- 吉村 仁, 2012: なぜ男は女より多く生まれるのか 絶滅回避の進化論. 筑摩書房, 東京, pp.47-70.

付録 統計学と確率論の歴史

筆者は現在, 連携大学院である東京海洋大学で集中講義を行っているが, 受講生の一人から「ある先生から確率を使うなら測度論を勉強するようにと言われたのですが, 良い本があったら教えてください」と質問されたので, とりあえず小針 (1973) を読むように回答した。水産分野において入門的な教科書は出版されているが (日本水産学会水産教育推進委員会, 2011; 西村, 2012), 測度論のような専門分野に踏み込んだ本は皆無のようである。この付録では統計学と確率論の歴史的事項について概略を説明する。確率および積分に関する基礎事項については以下の専門書を推薦する。測度論やルベグ積分については伊藤 (1963) が定評のある教科書であるが, ハードルがかなり高いので, 最初に志賀 (1990) を読むのがよい。確率微分方程式についての最近の教科書として西山 (2011) や森 (2012) がある。とりわけ森 (2012) は初心者向けに分かりやすく書かれている。

統計学の歴史

歴史的な事項について概略を述べると, 古典的な確率論はラプラスによって纏められた (Todhunter, 1865)。その後, ルジャンドルやガウスによって最小2乗法が発見されたが (安藤, 1995), 天文学における誤差解析が主流であった。近代的な統計学はゴルトンからで, チャールズ・ダーウィンの従兄弟であったため,

ダーウィンから生物データの解析を依頼され、回帰分析や相関、信頼楕円などを開発した。これを引き継いだのがカール・ピアソンである。(蓑谷, 1985; 安藤, 1997)。

ピアソンは応用数学の教授であったが、若い頃に「科学の文法」という本を書き、1900年にメンデル則が再発見されると、遺伝学の推進役を担った。しかし数学を生物学に応用するには相当な抵抗があったため、ゴルトンの援助を受け、専門の雑誌「バイオメトリカ」を創刊した。またコンピューターのない当時、多くの優秀な部下を率いて緊急の要であった重要な数表を作成した(蓑谷, 1987)。

ピアソンの弟子の一人であるゴセット(ステューデント)が t 分布を発見したが、この重要性に気づいたのはフィッシャーであった。それまでの大標本を用いる統計学から小標本の統計学に脱皮させたのがフィッシャーであるが、ピアソンの配下に入るのを嫌ったため、両者の関係は次第に悪化していった。カイ二乗適合度検定などで論争したが、数学的な問題についてはフィッシャーの方が正しかった。ピアソンが自由度を間違えたのは、信じられないミスであるが、非常に多忙であったことも一因だろう(安藤, 1989)。

ネイマンはポーランド出身で、若い頃にルベグ積分の新しい定理を発見するなど数学者として一流であった。シェルピンスキーらによってピアソンの所に留学させられたが、一年後パリに移ってボレルやルベグに学んでいる。カール・ピアソンの息子のエゴン・ピアソンはゴセットやフィッシャーの影響を受けたが、数学的能力に優れたピアソンに共同研究を依頼し、この共同研究によってネイマン・ピアソン流の検定法などが生まれた。信頼区間などはもともとフィッシャーのアイデアだったらしいが、実験計画法の解釈などでネイマンと対立したフィッシャーがfiducial 確率を唱えたため、帰無仮説を用いる方法はネイマン・ピアソン流と呼ばれている(Reid, 1982; 蓑谷, 1988; 竹内・大橋, 1981)。なおfiducial 確率とはパラメータ値の「確からしさ」の分布としての「確率」のことで、パラメータがfiducial 区間の中に存在するという命題の信頼性を表す尺度とされている(蓑谷, 1988; 竹内・大橋, 1981)。

カール・ピアソンは応用数学の教授であったから、数学的な技量は一流であったが、基礎的な部分にやや弱かったようである。ネイマンの論文について議論した際、ネイマンの方が正しかったが、「それはポーランドでは真理かもしれないが、ネイマン君、ここでは真理じゃない!」と怒って部屋から出て行ったという逸話がある(Reid, 1982)。一方、フィッシャーは幾何学的

な直感力に優れ、 n 次元の球帯の面積なども計算している(安藤・門脇, 1981; 安藤, 1997)。なおゴセットは「3次元以上(の数学)には精通していません」とカール・ピアソン宛の手紙に書いている(安藤・門脇, 1980)。

フィッシャーがネオ・ベイジアン(ベイズ)の筆頭であるサヴェジに「統計屋というのはあまり数学を振り回してはいかん。数学にあまり通じてはいかん」と言ったという話(竹内・大橋, 1981)は、ネイマンへの当てこすりと思われる。なおフィッシャーは集団遺伝学やゲーム理論における功績も大きい。日本生態学会(2004)でも重要な数理生態学者として紹介されており、またその中のハミルトンはフィッシャーを学問上の師と仰いでいる。

ネイマンがロンドンに派遣される際の逸話は、当時の状況がうかがえて興味深いので、ネイマンの言葉を以下に引用する。「変り者の二人、シェルピンスキーとバツサリーク、がある日やって来て、私に言いました。「知っているかね。この統計的研究集団がやっていることすべてを。そしてポーランドではそれが良いのか悪いのか誰も分からないのだよ。それで私たちはカール・ピアソンと一緒に研究するために一年間ロンドンへ君を派遣しようと思う。というのは、ピアソンは世界最大の統計学者だからね。それでもピアソンの雑誌、つまり「バイオメトリカ」だね、に君の何か論文を彼が発表してくれたなら、君は合格したと、私は思うことにしよう。もしそうならなかったら、帰ってこなくてよい」(Reid, 1982)。

確率論の歴史

一般に確率論は数学なので難しく、統計学は応用数学なので易しいという印象がある。確率論の歴史についてLaplace(1812)の付録に伊藤清が、伊藤(2004)の付録に池田信行が概略を述べている。最初に大数の弱法則をヤコブ・ベルヌイが示した。中心極限定理の原型は1718年にド・モアブルが導き、1812年にラプラスが一般化した。ただしカール・ピアソンはド・モアブルが $p \neq 0$ の場合も証明抜きで導いていることを指摘した(蓑谷, 1987)。古典的確率論はLaplace(1812)で完成したが、この本は特性関数を用いているので難解である(冗長な部分もあるが、結論はすべて正しい)。なおベイズの定理を一般化したのもラプラスであるが、Laplace(1814)では「太陽が明日もまた昇ることについては1に対する1826214の賭率を与えることができる」などと書いている。

1902年にルベグ積分が提示され、1909年にボレル

が大数の強法則を証明した。連続時間の確率過程はシュバリエやアインシュタインが提示し、1923年にウィナーが定式化した。1933年に「コルモゴロフの公理」が提示されて確率論が確立された (Kolmogoroff, 1974)。これに対しミーゼスは「確率の頻度説」を提示している。ただし吉田 (1977) によるとミーゼスの定義は国語辞典における言葉の説明のようなもので、中心極限定理のような法則を導くことができないが、コルモゴロフの公理からはすべての法則を導くことができる。コルモゴロフの公理は排反事象 A_1 と A_2 について、

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

という確率に要求されるべき基本性質が、

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

となるように理想化したものである。このような数学モデルがルベーグ測度として既に開拓されていたので、コルモゴロフはこれを矛盾のない公理として提出できたわけである。

大数の弱法則は

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{r}{n} - p\right| = 0\right) = 1 \quad (\text{A.1})$$

と表され、強法則は

$$P\left(\lim_{n \rightarrow \infty} \left|\frac{r}{n} - p\right| = 0\right) = 1 \quad (\text{A.2})$$

と表される (Geiringer, 1968)。前者は「確率収束」なので、若干の例外 (範囲から「はみだす」もの) があってもよいが、後者ではそのような例外は存在しない。極限操作 \lim の位置を確率 P の外から中に入れるのに200年近くかかったわけである。

コルモゴロフの公理を高木 (2010) が絶賛しているが、測度論に立脚しているので本当の価値を理解するのは難しい。大数の強法則はこれも測度論に立脚しているので証明は難しい。ルベーグはボレルの弟子であったが、二人の仲は良くなかった。ルベーグ積分の先取権のせいらしい (Reid, 1982; 好田, 1984a,b)。なおコルモゴロフや当時のソ連の状況については Gessen (2009) が詳しい。

(和文要旨)

負の二項分布と対数二項分布について経済学や生態学の分野で数学的な誤りが散見される。これらの修正、最高確率密度区間および確率一致事前分布を用いたベイズ統計手法について解説する。また枠どり法、超幾何分布、ランダム・ウォークモデル、ガンマ関数とベータ関数、カイ二乗分布について解説する。これらは水産科学における統計学手法のバックグラウンドであり、これらの知識は水産資源研究において役に立つ。