まえがき

本書は，筆者の大学での講義ノートを基に，教科書用として再構成したものである。

筆者は理学部物理の出身で，現在，乱流現象に対する理論と数値シミュレーションによる研究を行っている。乱流現象は古典物理学における未解決なランダム現象である。その流れ自体はナビエ・ストークス方程式で記述されており決定論的な問題であるが，非線形偏微分方程式の一般解法が未構築な現時点においては，そのカオティックな挙動から確率論的なアプローチが必要不可欠なものとなっている。そのため，私にとっては確率論や統計処理は非常に身近なものであるが，純粋数学的証明よりも物理現象の理解につなげる学問的応用におもな興味がある。そのため，本書では証明も示したが，物理などでの確率論の応用も検討している。

また，講義を行っている工学部は伝統的にあまり数学的な教養を重視しない分野であるが，実験計測や観測の際のデータ解析において，最近は確率論や統計学の必要性も高まってきている。確率や統計のいくつかの公式を記憶して使えるようになることがその際に重要であろうが，ただ公式に使われるといった勉強の仕方ではない何かがないかと筆者自身日頃心がけて講義してきた。その一端でもこの本に反映できればと，乱数シミュレーションなどに関する解説を本書において多めに割いたので，読者の一部には実際にトライしてより深い理解につながればと考えている。

最後にこの本の執筆を勧めていただいたコロナ社に心から感謝申し上げる。

2013 年 7 月

岡本　正芳

目　　　次

1.　確率論の基礎

2.　統計量の基礎

3.　離　散　分　布

4.　連　続　分　布

5.　正　規　分　布

6.　母集団と標本

7.　標　本　分　布

8. 検定・区間推定

9. 乱数シミュレーション

1 確率論の基礎

確率論は数学の一分野であるが，その数学的議論の起源は 17 世紀のパスカルとフェルマーによる，シュバリエ・ド・メレのギャンブルに関連した問題「4 回までサイコロを振って 1 の目を出す確率はいくつか。24 回までサイコロを 2 個同時に振って，$(1,1)$ の目を出す確率はいくつか。」にあるともいわれている。歴史的にはその後，少し数学や物理学を勉強した人なら耳にするであろうベルヌーイ，ラグランジュ，ポアソン，ラプラスにより古典確率論が確立された。また，ガウスは統計処理の先駆けともなる著書「誤差論」で最小 2 乗法や正規分布などを明示してきた。20 世紀になると，コルモゴロフが測度論を導入して現代確率論を構築し，ウィーナーとレヴィによる確率過程論，伊藤清による確率解析学へと発展してきた。余談であるが，現代確率論の父ともいうべきコルモゴロフは，著者の研究対象である “乱流” においても，ランダムな現象の中に隠れた普遍則であるコルモゴロフスペクトルの発見という，偉大な研究成果を挙げており，個人的には親近感がわく人物である。

一方，物理学分野では 17 世紀にニュートンによって提唱された力学により，不確かさの入り込む余地のない決定論的世界観が構築された。それに対し，19 世紀にマクスウェルやボルツマンによる熱力学を，アボガドロ数レベルの多数の分子の力学運動の集合体として捉え直す統計力学の構築によって，確率論の導入が大きな成果をおさめた。これは，物理学における基盤数学の拡張が果たされたといえる。その後，20 世紀になると量子力学の基礎原理であるハイゼンベルグの不確定性原理「粒子の位置と運動量，エネルギーと時間などの一組みの物理量について，両者を同時に正確に測定し，決定することはできない。」に基づき，確率論的世界観が我々の根源に存在することが明らかとなった。これらの点から，物理学において確率論は利用云々（うんぬん）とは違って，基盤となるべき重

要な数学分野である。

さらに近年の確率・統計分野の進展は目覚ましく，数学や物理学といった自然科学のみならず，より実用的な工学や経済学など様々な利用が行われるようになってきた。元々，確率論はギャンブルなどと関連した実用性の高い数学分野であり，いろいろな研究分野に容易に適用することができ，普段の実験や数値シミュレーションなどで生じるデータ解析の基盤となるものでもある。よって，学生諸氏にとっては確率・統計を勉強することは有意義なものになるであろう。

1.1　確率変数と確率関数

確率論における目的は，確率現象によって生じる事象がどれくらいの確率で生じるかを判断することである。そこで，事象を数学的に取り扱うため，それに割り当てる**確率変数**（random variable）を設定する必要がある。例えば，コインを投下して生じる事象は 2 通りあり，表が出るか，裏が出るかである。これに対して確率変数 x を表を 0，裏を 1 と設定すると，表が出る確率は 1/2，裏が出る確率も 1/2 となる。この確率を確率変数 x を用いて関数表現化したものが**確率関数**（probability function）$p(x)$ であり

$$p\,(x) = \begin{cases} \dfrac{1}{2} & x = 0, 1 \\ 0 & x \neq 0, 1 \end{cases} \tag{1.1}$$

となる。この確率関数は **図 1.1**(a) のようなグラフとなる。確率関数は確率自体を表している。確率変数は表を 1，裏を 2 とおけば，その確率関数の結果は変更しなければならない。また，サイコロであれば，生じる事象は 1 の目，2 の目，3 の目，4 の目，5 の目，6 の目のどれかの 6 パターンである。この場合，出目の値をもって確率変数 x を設定すれば，図 1.1(b) のようになる確率関数は

$$p\,(x) = \begin{cases} \dfrac{1}{6} & x = 1, 2, 3, 4, 5, 6 \\ 0 & x \neq 1, 2, 3, 4, 5, 6 \end{cases} \tag{1.2}$$

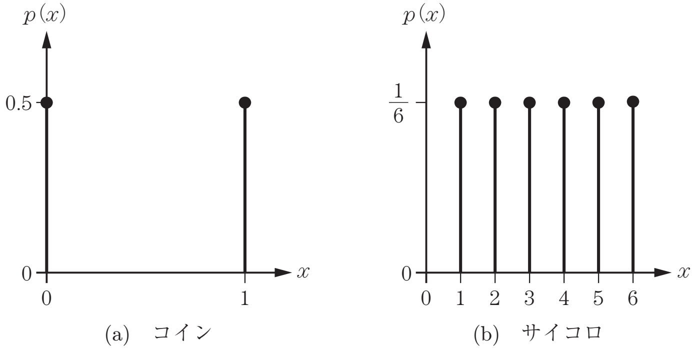


図 1.1　コインとサイコロの確率分布

である。これら確率変数と確率関数は確率現象を数学的に記述しており，**確率分布**（probability distribution）を構成している。これらの例は確率変数 x が有限な個数の自由度（選択肢）しかなく，せいぜい可算無限個の自由度の確率分布と合わせて，**離散分布**（discrete probability distribution）と呼ばれる。

確率関数 $p(x)$ に課せられた制約は，以下の 2 点である確率値の正値性と全確率が 1 となることである。

$$p(x) \geqq 0 \tag{1.3}$$

$$\sum_{x}^{\mathrm{All}} p(x) = 1 \tag{1.4}$$

$p(x)$ を x を引き数とする正実数の並びと解釈すると，この性質を持っている有限および無限級数列はすべて確率関数として利用できる。

つぎに例えば，身長といったものに着目してみよう。当然いろいろな人がいるように身長は個々人それぞれ違っており，様々な値をとりうる量となっている。このような量を確率変数 x と設定すると，x は連続なある範囲の値をとる。この場合，ある値になる確率には本質的に意味がない。非可算無限個の離散的な自由度がある離散分布と考えれば，特定の値になる確率はおよそ $1/\infty = 0$ であるからである。このような確率変数が連続な値をとる確率分布は**連続分布**（continuous probability distribution）と呼ばれる。連続分布では確率関数

$p(x)$ を議論することは妥当性がなく，確率変数 x がある範囲内の値をとる確率を議論することが有効である。そこで，以下の二つの関数が対象となる。

一つは**累積分布関数**（cummulative distribution function）$F(x)$ と呼ばれる関数で，確率変数 x' がある値 x 以下の値をとる確率を意味する。

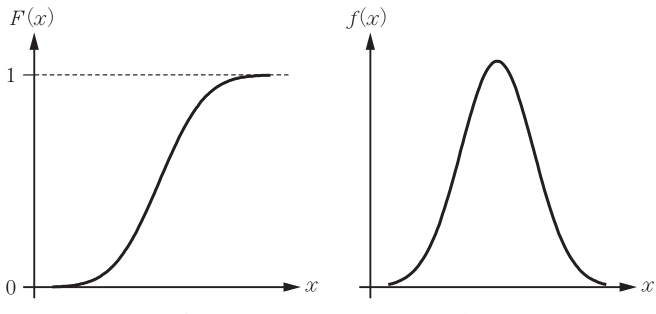
$$F(x) = p(-\infty \leqq x' \leqq x) \tag{1.5}$$

累積分布関数は確率を表しており，確率に対する正定値と全確率から，$F(x)$ に対する制約は

$$0 \leqq F(x) \leqq 1 \tag{1.6}$$

$$F(\infty) = 1 \tag{1.7}$$

となる。そして，$F(x)$ は図 **1.2**(a) のような単調増加の関数である。



(a)　累積分布関数　　(b)　確率密度関数

図 1.2　連続分布の例

もう一つの関数は**確率密度関数**（probability density function）$f(x)$ で累積分布関数 $F(x)$ の確率変数 x に関する導関数

$$f(x) = \frac{dF(x)}{dx} \tag{1.8}$$

で定義される。この関数値は直接は確率を意味せず，確率変数 x が有次元量である場合，その逆数の次元を持っている。簡略表現としては “PDF” と記述される

ことも多々あるので覚えておくとよいであろう。この関数の名称に付いている“密度” は積分すると確率を与えることを意味しており，確率変数が $a \leqq x \leqq b$ ($b > a$) の範囲にある確率は

$$p\,(a \leqq x \leqq b) = F\,(b) - F\,(a) = \int_a^b dx f\,(x) \tag{1.9}$$

によって求まる。確率密度関数の正定値と全確率の制約は

$$f\,(x) \geqq 0 \tag{1.10}$$

$$\int_{-\infty}^{\infty} dx f\,(x) = 1 \tag{1.11}$$

となる。確率密度関数 $f(x)$ が図 1.2(b) のような場合，ピーク付近での事象が発生しやすいと考えてよい。この本の中では，主として後者である確率密度関数を用いて連続分布を解説していく。

1.2 事 象 の 独 立

二つの確率事象 A と B があるとする。事象 A が生じると仮定したとき，事象 B が生じる確率を**条件付き確率**（conditional probability）$p\,(B\,|A\,)$ と呼び，それは

$$p\,(B\,|A\,) = \frac{p\,(A \cap B)}{p\,(A)} \tag{1.12}$$

で与えられる。ここで，$A \cap B$ は A と B の積事象（共通集合）で，事象 A と B がともに生じることを意味している。つまり，A と B がともに生じる確率を A が生じる確率で除すことで条件付き確率が導出される。例えば，サイコロを投げる場合において，事象 A は「偶数の目が出ること」，事象 B は「3 より小さい目が出ること」とすると，事象 $A \cap B$ は「2 の目が出ること」だけとなる。個々の確率は $p\,(A) = 1/2$ と $p\,(A \cap B) = 1/6$ であることから，条件付き確率は $p\,(B\,|A\,) = 1/3$ となる。

事象 A と B があって，事象 A が生じる生じないが事象 B が生じることに影

響を及ぼさないとき，事象 A と B は**独立**（independence）であるという。つまり，事象 A が生じるという仮定そのものが意味を持たないので

$$p\,(B\,|A\,) = p\,(B) \tag{1.13}$$

ということが，独立の条件式になる。条件付き確率の定義式 (1.12) と上式から事象の独立は次式でも与えられる。

$$p\,(A \cap B) = p\,(A)\,p\,(B) \tag{1.14}$$

左辺の $p\,(A \cap B)$ の確率を**結合確率**（joint probability）と呼び，$p\,(A, B)$ とも表記される。このように二つの確率現象から構成される場合，2 次元確率分布を議論する必要がある。独立が成立しているとき，事象 A と B の確率関数の単純な積で 2 次元確率関数を表すことができる。例えば，サイコロを投げることとコインを投げることを考えよう。前者に確率変数 x を，後者に確率変数 y を設定すると，それらのとりうる値は

$$x = \{1, 2, 3, 4, 5, 6\} \tag{1.15}$$

$$y = \{0, 1\} \tag{1.16}$$

とすることができ，2 次元確率変数 (x, y) としては

$$\begin{aligned}(x, y) = \{&(1, 0)\,, (2, 0)\,, (3, 0)\,, (4, 0)\,, (5, 0)\,, (6, 0)\,,\\ &(1, 1)\,, (2, 1)\,, (3, 1)\,, (4, 1)\,, (5, 1)\,, (6, 1)\}\end{aligned} \tag{1.17}$$

の 12 パターンとなる。両試行には関連性がないので独立が仮定できる。サイコロに関しては偶数の目が出て（$x = 2, 4, 6$），コインは表が出る（$y = 0$）といった事象が生じる確率 $p(x, y)$ は $(1/2) \times (1/2) = 1/4$ となり，12 パターン中の 3 パターンが該当する確率 $3/12 = 1/4$ と一致する。また，この独立を連続分布に拡張しておこう。一つの確率分布（確率変数 x_1）は累積分布関数 $F_1(x_1)$，確率密度関数 $f_1(x_1)$ で，もう一つは確率変数 x_2 で累積分布関数 $F_2(x_2)$，確率密度関数 $f_2(x_2)$ とする。考慮する範囲 $(a_1 \leqq x_1 \leqq b_1, a_2 \leqq x_2 \leqq b_2)$ の確率

$p_{1,2}$ は以下のように変換できる。

$$
\begin{aligned}
&p_{1,2}\left(a_1 \leqq x_1 \leqq b_1,\ a_2 \leqq x_2 \leqq b_2\right)\\
&\quad = p_1\left(a_1 \leqq x_1 \leqq b_1\right) p_2\left(a_2 \leqq x_2 \leqq b_2\right)\\
&\quad = \left(F_1\left(b_1\right) - F_1\left(a_1\right)\right)\left(F_2\left(b_2\right) - F_2\left(a_2\right)\right)\\
&\quad = \int_{a_1}^{b_1} dx_1 f_1\left(x_1\right) \times \int_{a_2}^{b_2} dx_2 f_2\left(x_2\right)\\
&\quad = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 f_1\left(x_1\right) f_2\left(x_2\right)\\
&\quad = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 f_{1,2}\left(x_1, x_2\right)
\end{aligned}
\tag{1.18}
$$

よって，独立が成立しているときの 2 次元連続分布の確率密度関数は

$$
f_{1,2}\left(x_1, x_2\right) = f_1\left(x_1\right) f_2\left(x_2\right) \tag{1.19}
$$

となり，確率関数と同じように個々の確率密度関数の積で求めることができる。この本では以降，この性質はたびたび利用するので覚えておいてほしい。

せっかく，条件付き確率を持ち出したので，**完全確率の公式**（law of total probability）と**ベイズの公式**（Bayes' theorem）を示しておく。n 個の事象 A_i（$i = 1, \cdots, n$）が互いに積事象がなく，つまり排反

$$
\bigcup_{i=1}^{n} A_i = \Omega \tag{1.20}
$$

とする。完全確率の公式は

$$
p\left(B\right) = \sum_{i=1}^{n} p\left(A_i\right) p\left(B \left| A_i\right.\right) \tag{1.21}
$$

となり，ベイズの公式

$$
p\left(A_k \left| B\right.\right) = \frac{p\left(A_k\right) p\left(B \left| A_k\right.\right)}{\sum\limits_{i=1}^{n} p\left(A_i\right) p\left(B \left| A_i\right.\right)} \tag{1.22}
$$

となる。これらの公式を以下の問題で具体的にどんなものか見てみよう。

パソコン 1 台には a 社，b 社，c 社が生産した部品をそれぞれ 10%，30%，

60%含んでいる。それぞれの会社は0.5%，0.4%，0.1%の確率で不良品が発生することがわかっていると仮定する。それぞれの事象を以下に与える。

事象 F：不良品である。

事象 A：a社の部品である。

事象 B：b社の部品である。

事象 C：c社の部品である。

パソコン1台の構成を考えると，全確率から

$$p(A) + p(B) + p(C) = 0.1 + 0.3 + 0.6 = 1 \tag{1.23}$$

となる。

まず，パソコンが不良品である確率 $p(F)$ を求めよう。組み立てられたパソコンが不良品であるということは，パソコン内のどれかの部品が不良品であることが原因であるから，確率は完全確率の公式より

$$\begin{aligned} p(F) &= p(F|A)\,p(A) + p(F|B)\,p(B) + p(F|C)\,p(C) \\ &= 0.005 \times 0.1 + 0.004 \times 0.3 + 0.001 \times 0.6 \\ &= 0.0005 + 0.0012 + 0.0006 = 0.0023 \end{aligned} \tag{1.24}$$

と求まり，不良品である確率は0.23%ということになる。完全確率の公式は，構成要素の情報から全体としてどのようになるかを判断するのに利用できる公式である。

つぎに不良品が判明したとき，その原因がc社の部品に帰する確率を求めよう。この場合はベイズの公式を利用して以下のように求める。

$$\begin{aligned} p(C|F) &= \frac{p(F|C)\,p(C)}{p(F|A)\,p(A) + p(F|B)\,p(B) + p(F|C)\,p(C)} \\ &= 0.001 \times 0.6 \div 0.0023 = 0.26086\cdots \end{aligned} \tag{1.25}$$

c社の部品が原因である確率はおよそ26%となり，c社の製品はパソコン内に多数含まれているが，不良品発生率の低さから1/3よりも低いものとなっている。ベイズの公式は完全確率の公式とは逆に，個々の構成要素に帰する情報を把握することに利用できる。

索　　　　　引

—— 著 者 略 歴 ——

1992 年　東京大学理学部物理学科卒業
1994 年　東京大学大学院理学系研究科修士課程修了（物理学専攻）
1997 年　東京大学大学院理学系研究科博士課程修了（物理学専攻）
　　　　博士（理学）
1997 年　静岡大学助手
2004 年　静岡大学助教授
2007 年　静岡大学准教授
　　　　現在に至る

工学系のための確率・統計
——確率論の基礎から確率シミュレーションへ——
Probability and Statistics for Students in Engineering Course
——From Fundamentals of Probability Theory to Stochastic Simulation——

2013 年 9 月 5 日　初版第 1 刷発行　★

検印省略

おか　もと　まさ　よし
著　　者　岡　本　正　芳
発 行 者　株式会社　コ ロ ナ 社
　　　　　代 表 者　牛 来 真 也
印 刷 所　三 美 印 刷 株 式 会 社

112–0011　東京都文京区千石 4–46–10
発行所　株式会社　コ　ロ　ナ　社
CORONA PUBLISHING CO., LTD.
Tokyo Japan
振替 00140–8–14844・電話(03)3941–3131(代)
ホームページ　http://www.coronasha.co.jp

ISBN 978–4–339–06104–8　（横尾）　（製本：愛千製本所）
Printed in Japan

落丁・乱丁本はお取替えいたします