第 5 章

正規分布

統計学の中心となる概念は正規分布である．正規分布はいたるところに現れる確率分布で，多くの場合にその性質を使った分析は役に立つ情報を与えてくれる．

5.1 離散的確率分布から連続的確率分布へ

5.1.1 実数データの取り扱い

これまでの確率分布はすべて離散的な確率変数をもつものとして取り扱ってきた．しかし，私たちが現実に取り扱うデータの多くは整数ではなく，小数点を含むものである．たとえば体重の分布を考えてみると，1 章で扱ったように，ふつうは 0.1 kg 刻みのデータとして収録されているであろう．もっと考えると，そもそもある人の体重は，

58.32417081... kg

のように，現実に測定はできないものの，限りなく細かいところまで決まっているはずである[*1]．すなわち，**多くの数値データは実数として存在している**のである．

すなわち，これまでは離散的な確率変数を扱ってきたのだが，我々が扱いたい数値データの多くは実数であるので，離散的な変数には乗らないのである．このような**連続的な確率分布** を本章では取り扱う．中でも重要なのは正規分布である．

5.1.2 離散的確率関数の形

離散型一様分布というのは，3.2 節で紹介したように，サイコロの目式の確率分布である．すなわち 1 から n までのいずれかの目が等しい確率 $1/n$ で出現するような確率分布のことである．ちなみにサイコロの目についての確率分布をグラフで表すと図 5.1 のようになるだろう．

[*1] 実際には，こんな細かい値は息をするだけで変動しているであろうが，それでもある一瞬一瞬では相当に細かいところまで確定しているに違いない．

この図の意味するところは，確率関数 $P(X)$ は，$X = 1, 2, ..., 6$ の点でだけ 1/6 という値をとり，それ以外では至るところゼロであるような，櫛の歯のような関数であるということである．サイコロの目というのは，1 から 6 の目を取るだけで，その間の中間的な値などは取りようもないのだから，これは当然である．このように**櫛の歯型の関数になるというのは，離散型確率関数の特徴である．**

5.1.3　連続領域での確率の特徴

離散型一様分布に対して**連続型一様分布**というのは，ある区間に落ちてくる雨粒の分布のようなものである．今，地面に 2 m の長さの線を引き，落ちてくる雨粒の位置が，その上のどこに来るかということを考えてみよう．雨は広い区域でまんべんなく降るのであるから，この区間の中のどこでも，雨粒が落ちてくる確率が等しいだろうということは，容易に想像できる．このようすを図にするならば，下の図 5.2 のようになると思われるであろう．しかし，この時，グラフの縦軸の値はどうすればいいのだろうか？

離散的な確率関数の場合，図 5.1 にあるように，縦軸の値が意味するのは，確率変数 X がある値をとるときの確率そのものであった．ところが，図 5.2 のように $[0, 2]$ の連続した区間[*2]にあるすべての実数について一定の確率 p が与えられているとすると，実数というのは有限の区間の中に無限に存在するのだから，それら無数に多くの点についてその確率が適用されることになり，全部の確率を加えた値はかならず無限大になってしまう[*3]．

5.1.4　連続分布の確率は面積で表現する

上述の問題を解決するには，連続的な確率変数の場合の確率の値を，グラフ上の面積で表すということにするとよい．図 5.3 は，長さ 2 m の区間に一様に雨が降ってくる確率

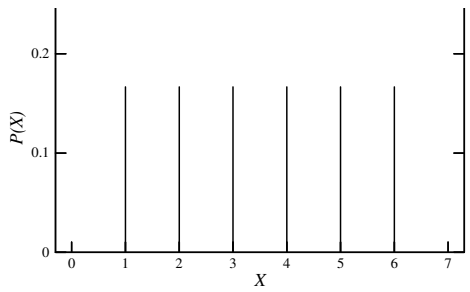

図 5.1　サイコロの目の出方の確率分布関数

[*2] たとえば 1 と 2 と，その間のすべての実数を含む区間のことを，$[1, 2]$ というふうに表す．

[*3] かといって，ある一点にちょうど雨粒が来る確率は限りなく小さいはずだからという理由で，p をゼロとしてしまうと，いくら加えてもゼロのままになって始末におえなくなる．

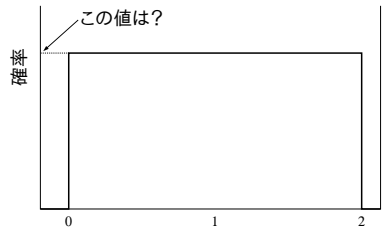

図 5.2 雨粒がある区間に落ちる確率は一様である．その確率の値はどう決めればよいのだろうか

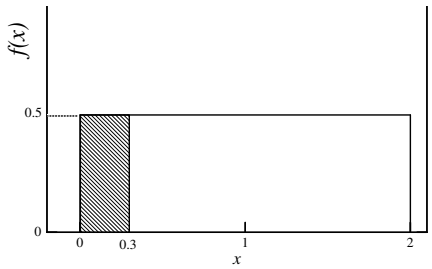

図 5.3 2 m の区間で定義された連続型一様分布の確率密度関数

を表す関数のグラフである．具体的に式で書くと次のようになる．

$$\begin{aligned} f(x) &= 0.5, \quad (0 \leq x \leq 2) \\ f(x) &= 0, \quad \text{それ以外} \end{aligned} \tag{5.1}$$

ここで注目しておかねばならないのは，グラフの縦軸の値が 0.5 になっていて，全面積を 1 にするように関数の値が決められているということである．

こうしておけば，この関数の定義域である $[0, 2]$ の区間のグラフの面積はちょうど 1 になる．確率は 1 の時に全事象を覆うのであるから，**面積を確率とみなすことで，連続な分布の時の確率をうまく表現できる**ことになる．したがって，たとえば図の左側の $[0, 0.3]$ の区間に雨が落ちる確率は，斜線で示された部分の面積，$0.3 \times 0.5 = 0.15$ というふうに表現される．

このことを数学的に表しておこう．区間 $[a, b]$ の中で，ある関数 $f(x)$ が描く面積は定積分で表されるから，x がその区間に収まる確率 $P(a \leq X \leq b)$ は，次のように書ける．

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx \tag{5.2}$$

また，起き得るあらゆる場合について確率を足し合わせると 1 になることから，

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{5.3}$$

となる[*4]．

この例に即して書けば，$f(x) = 0.5,\ a = 0,\ b = 0.3$ であるから，

$$P(0 \leq X \leq 0.3) = \int_{0}^{0.3} 0.5\,dx = [0.5x]_{0}^{0.3} = 0.15 \tag{5.4}$$

となる．

このように，連続型の確率変数 x に対して，確率の大きさを表す関数 $f(x)$ を**確率密度関数 (probability density function)** という．

また，$f(x)$ の**累積分布関数** $\Phi(z)$ は次のように与えられる関数である．

$$\Phi(z) = \int_{-\infty}^{z} f(x)\ dx \tag{5.5}$$

これは，図 5.4 を見れば分かるように，$f(x)$ のあるところまでの面積で表される値であり，確率の定義から，

$$\Phi(-\infty) = 0 \tag{5.6}$$

$$\Phi(\infty) = 1 \tag{5.7}$$

となっていなければならない．

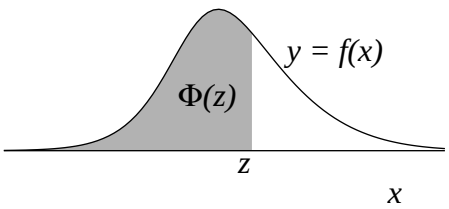

図 5.4　累積分布関数 $\Phi(z)$ は，ある範囲の事象を覆う確率密度関数 $f(x)$ の面積で定義され，その範囲の事象が起きる確率を表現する．

図 5.3，式 5.1 で示される確率密度関数については，その累積分布関数は

$$\Phi(z) = \int_{-\infty}^{z} 0.5\ dx = 0.5z, \quad (0 \leq z \leq 2) \tag{5.8}$$

[*4] ここで積分範囲を $[-\infty, \infty]$ にとったのは，X のすべての範囲について積分するという理由からである．しかし，実際には確率が定義されている範囲にわたって積分を計算すればよい．

と表せる．

例題 5–1 式 5.1 で示される確率密度関数を使って，一滴の雨粒が端から測って 0.8 m の点から 1.5 m の点の間に落ちる確率を求めよ．

$f(x) = 1/2$ を単に $x = 0.8$ から $x = 1.5$ まで積分するだけである．従って，

$$\int_{0.8}^{1.5} \frac{1}{2}\, dx = \left[\frac{x}{2}\right]_{0.8}^{1.5} = 0.35$$

もちろん図 5.3 を見て，相当する範囲の面積を求めるだけでもよい．

5.1.5 連続的確率関数の平均，分散

ある確率密度関数が与えられたとき，その平均や分散がどうなるかを導いておこう．その前に，離散型の場合の式 (3.6) に相当する関係式として，

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{5.9}$$

という条件が満たされている必要がある．この条件は，全事象に対する確率の和が 1 であることを意味しており，**規格化条件 (normalizing condition)** という．

さて, 期待値 (平均) を表すための式 (3.10) に相当するのは，次の式である．

$$E[X] = \mu = \int_{-\infty}^{\infty} x f(x)\, dx \tag{5.10}$$

さらに，分散を表す式 (3.11) に相当するのは次の式である．

$$V[X] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx \tag{5.11}$$

これらは単に 3.4 節の関係式における総和を積分に変えただけのものである．

式 (3.12) はここでも成立している．すなわち，

$$V[X] = E[X^2] - \{E[X]\}^2 \tag{5.12}$$

ただし，

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx \tag{5.13}$$

である．

5.1.6　コンピュータで発生させる一様乱数

コンピュータを使うと区間 $[0,1]$ で一様な確率密度をもつ乱数を発生させることが簡単にできる．たとえば，下は試しに 10 個の乱数を発生させてみたものである．

0.2747, 0.2288, 0.6893, 0.1855, 0.9086, 0.1876, 0.9291, 0.5324, 0.3335, 0.8568

この乱数は式で表すと次の一様連続分布に従っている．

$$\begin{aligned} f(x) &= 1, \ (0 \leq x \leq 1) \\ f(x) &= 0, \ (\text{それ以外}) \end{aligned} \tag{5.14}$$

この分布の期待値が $1/2$ であることはほとんど自明だが，式 (5.10) を適用してみると，

$$\mu = \int_0^1 x \times 1 \ dx = \frac{1}{2} \tag{5.15}$$

となって，その通りになっている．

一方，分散は式 (5.12) より，

$$\sigma^2 = \int_0^1 x^2 \times 1 \ dx - \mu^2 = \frac{1}{12} \tag{5.16}$$

となる．このことについては章末の問題で触れることにする．

コンピュータを使うと，他にも，ある整数よりも小さい正の整数をランダムに発生させたり，一様でない分布に従う乱数を発生させたりすることが容易にできるようになっている．

5.2　二項分布から正規分布へ

二項分布 $B[n,p]$ で，n が非常に大きくなるとどのような分布を描くかを調べてみよう．図 5.5 は，$p=0.4$ として，n を変えてみた時に，二項分布の形がどのようになるかを調べたものである．

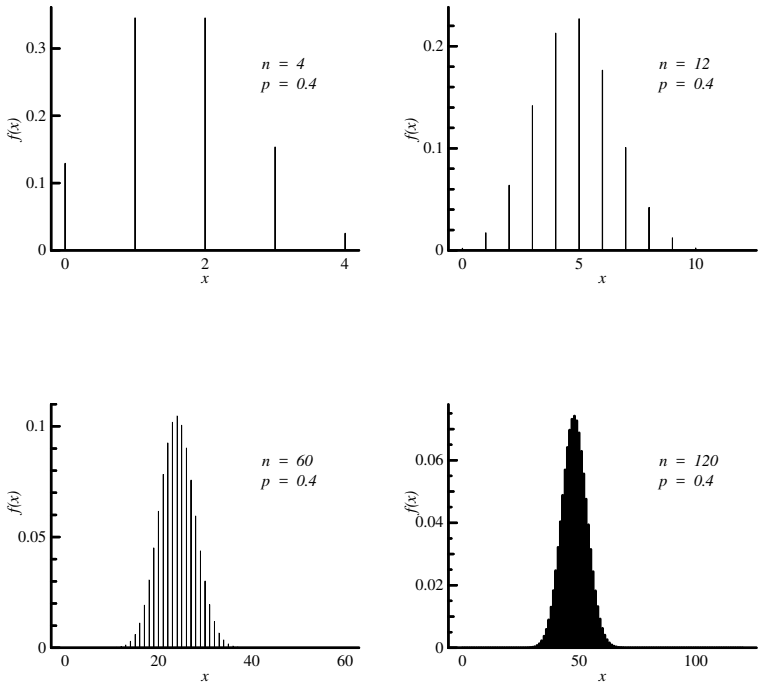

図 5.5　二項分布 $B[n,p]$ の n を増加させていったときの分布の形の変化

このグラフを見ると，二項分布の n が大きくなるに従って，平均値の周りに左右対称な吊鐘形をした分布になっていく．この分布は**正規分布 (normal distribution)** と呼び，確率統計において中心的な役割を果たす確率分布である．

正規分布は，次のような式で表される

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{5.17}$$

ただし，$\exp(a)$ は e^a と同じである[*5]．

ここで μ, σ^2 は，それぞれ正規分布の平均と分散である[*6]．

4.1.2 で見たように，二項分布においては $\mu = np$, $\sigma^2 = np(1-p)$ であるから，上述の二項分布で n を大きくしていった時の正規分布への移行は，

$$B[n,p] = {}_nC_x p^x (1-p)^{n-x} \overset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{5.18}$$

を意味する．

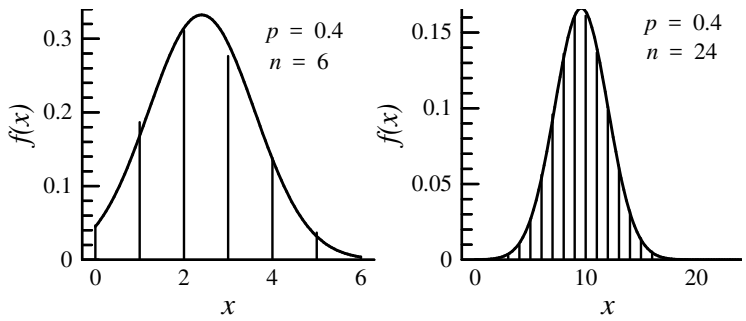

図 5.6　同じ σ と μ をもつ二項分布 (縦の線) と正規分布（実線）を $n = 6, 24$ に対して描いたもの．

この式の導出はやや手順が長いので省略するが，実際に二項分布と正規分布のグラフを重ねてみて，両者がどのように近づくかを見てみよう（図 5.6）．この図を見ると $n = 24$ の二項分布は正規分布ときわめてよく一致していることがわかる．実用上は，$np > 5$ かつ $n(1-p) > 5$ であれば，二項分布を正規分布として扱うことは差し支えないとされるので，$p = 0.5$ であれば $n = 10$ 程度でもよい近似が得られると考えてよい．

正規分布は，後にみるように非常に多くのケースに使われるもっとも重要な確率分布であり，しばしば $N[\mu, \sigma^2]$ と表される．すなわち，

平均 $\boldsymbol{\mu}$, 分散 $\boldsymbol{\sigma^2}$ をもつ正規分布を $\boldsymbol{N[\mu, \sigma^2]}$ と表す．

[*5] exp を使ったほうが，式がかさばらないので，しばしば使われる．

[*6] 正規分布の平均が μ であることは，式 (5.17) の形からすぐに分かる．すなわち，この関数は $x = \mu$ のときに最大値をとり，かつそこを中心に左右対称になっているわけであるから，平均は $x = \mu$ のところになっていなければならない．

5.3　正規分布表の活用

5.3.1　標準正規分布と標準化変換

平均が μ, 分散が σ^2 であるような確率変数 x が正規分布に従うものとしよう．このとき，式 (5.19) で定義される変換をほどこしてみる．

$$z = \frac{x-\mu}{\sigma} \tag{5.19}$$

すると z は式 (5.20) の分布に従う。

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{5.20}$$

これは平均がゼロで標準偏差（分散）が 1 であるような正規分布であり，**標準正規分布**と呼ばれる．また，式 (5.19) で定義される変換を**標準化変換 (standardization)** という．

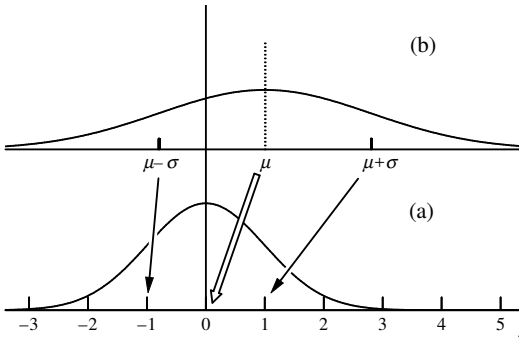

図 5.7　一般の正規分布 $N[\mu, \sigma^2]$ を $z = (x-\mu)/\sigma$ で標準正規分布 $N[0,1]$ に変換する．

図 5.7 に標準化変換の意味を示した．一目でわかるように，一般の正規分布 $N[\mu, \sigma^2]$ に従う確率変数 x が与えられたとすると，それを標準化変換して得られる z は正規分布 $N[0,1]$ に従う．

また，標準正規分布から一般の正規分布への逆の変換も考えられる．これは式 (5.19) からただちに導かれる．

$$x = \sigma z + \mu \tag{5.21}$$

一般の正規分布に従う分布を標準正規分布に変換することで，次に出てくる正規分布表を使ってさまざまな計算を進めることが可能になる．

表 5.1　正規分布表からの抜粋

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.500000	1.00	0.841345	2.00	0.977250	3.00	0.998650
0.10	0.539828	1.10	0.864334	2.10	0.982136	3.10	0.999032
0.20	0.579260	1.20	0.884930	2.20	0.986097	3.20	0.999313
0.30	0.617911	1.30	0.903200	2.30	0.989276	3.30	0.999517
0.40	0.655422	1.40	0.919243	2.40	0.991802	3.40	0.999663
0.50	0.691462	1.50	0.933193	2.50	0.993790	3.50	0.999767
0.60	0.725747	1.60	0.945201	2.60	0.995339	3.60	0.999841
0.70	0.758036	1.70	0.955435	2.70	0.996533	3.70	0.999892
0.80	0.788145	1.80	0.964070	2.80	0.997445	3.80	0.999928
0.90	0.815940	1.90	0.971283	2.90	0.998134	3.90	0.999952

5.3.2　正規分布表とその意味

正規分布を利用するさいには，**正規分布表**と呼ばれる数表を用いる．この表には，z に対して次の定積分の値 $\Phi(z)$ を計算したものを掲載してある．表 5.1 に抜粋した表を示す．

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \tag{5.22}$$

この積分は，図 5.8 の影をつけて示されている領域の面積に相当するので，標準正規分布に従う確率変数が $x = z$ よりも左側の部分に入る確率を意味することになる．

なお，ここで注意しておいてほしいのは，世の中の正規分布表には積分区間を $[-\infty, z]$ ではなく，$[0, z]$ としたものもあるということである．その場合には $z = 0$ に対して $\Phi(z) = 0$ となり，表の値は 0.5 だけずれることになる．ただし図を見て考えれば，このことは大して面倒なことではない．

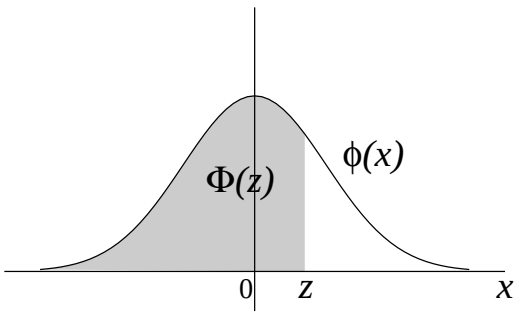

図 5.8　正規分布 $\phi(x) = 1/\sqrt{2\pi}\, \exp(-x^2/2)$ と累積分布関数 $\Phi(z)$ との関係

5.3.3 正規分布の計算の手順

標準正規分布関数 $f(z)$ について，常に次の 2 点を押さえておこう．

- $f(z)$ を $-\infty$ から ∞ まで積分して足し合わせた全面積は 1 である．
- $f(z)$ は左右対称である．

■標準正規分布における確率を求める

標準正規分布 $N[0,1]$ において，$-1 \leq z \leq 1$ の範囲に含まれる面積はどれだけか．

$z < 1$ の部分の面積は，表から 0.841345 である．したがって $z > 1$ に含まれる面積は $1 - 0.841345 = 0.158655$ であり，図 5.8 を見て考えれば，$z < -1$ に含まれる面積もそれに等しいことが分かる[*7]．よって答えは，$1 - 0.158655 \times 2 = 0.68269$

■標準化変換を使う

標準正規分布に従うような事象はあまりなく，一般的な $N[\mu, \sigma^2]$ に従う分布がほとんどであるから，よくある解き方の手順は次のような感じになる．

1. $x \to z$ の標準化変換を行う
2. 正規分布表から z よりも小さい部分の面積を読み取る
3. その値と，上記の 2 つの点を使って，求めるべき部分の面積を計算して確率とする。

例題 5–2 受験者の平均点が 65 点，標準偏差が 12.5 であるような試験があったとする．得点分布が正規分布しているとした場合，この試験で 90 点以上を取る人の割合は，全体のうち 何 % と見ればよいか．

式 (5.19) の変換によって，この成績での 90 点を標準化すると，$z = (90-65)/12.5 = 2.0$ になる．そこで $\Phi(2.0)$ を表から調べると，0.97725 であるから，100 人のうち 97.7 人が 90 点未満のところにいるとしてよい．従って答えは 2.3 % となる．

■半整数補正—二項分布を正規分布で近似する

5.2 節で扱ったように，二項分布は n がそこそこ大きければ正規分布に近似できる．そのため，面倒な組み合わせの計算をすることなく，数表と簡単な計算だけで二項分布に従う事象の確率を計算でき，非常に強力な確率計算の武器になる．

[*7] 図を見てちょっと考えれば，$2\Phi(z) - 1$ が求める答えであることが分かる．それを使うと計算はもっと簡単である．

ただし，離散的な確率分布を連続分布に置き換えて計算する都合上，**半整数補正**と呼ばれる近似手法がよく用いられる．次の例題でそのことを説明しよう．

例題 5–3　日本人で A 型の血液をもつ人は 40% いる．10 人の日本人を集めた時に，A 型の血液を持つ人が 4 人から 6 人の間になる確率を求めて，二項分布による計算の結果と比較せよ．

二項分布では $\mu = np$, $\sigma = \sqrt{np(1-p)}$ である．したがってこの集団については，$\mu = 4$, $\sigma = \sqrt{10 \cdot 0.4 \cdot 0.6} = 1.55$ となる．一方**この二項分布を，連続分布である正規分布とみなすと，4 人から 6 人の間であるということは 3.5 人と 6.5 人の間にあるとみなせる**から，$z_1 = (3.5-4)/1.55 = -0.32$, $z_2 = (6.5-4)/1.55 = 1.61$ の間に入る確率 $\Phi(1.61) - \Phi(-0.32)$ を計算すればよい．図 5.9 にそのことを示した．ここで負の z に対する値は正規分布表にないが，正規分布の形を考えれば $\Phi(-0.32) = 1 - \Phi(0.32) = 0.374$ であり，$\Phi(1.61) = 0.946$ なので，答は 0.572 となる．

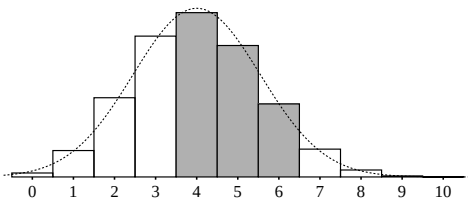

図 5.9　半整数補正の使い方．二項分布 $B[10, 0.4]$ で $x = 4, 5, 6$ に相当する積分範囲は 3.5〜6.5 になる．

▼半整数補正を使う条件　二項分布 $B[n, p]$ が与えられていた時，$[x_1, x_2]$ という区間の面積を求めたい．ここで x_1, x_2 は整数である．このとき半整数補正を使うかどうかを判断する基準を考えておこう．

段取りを見直すと，まず次のように x_1, x_2 を標準化する．

$$\begin{aligned} z_1 &= \frac{x_1 - \mu}{\sigma} \\ z_2 &= \frac{x_2 - \mu}{\sigma} \end{aligned} \tag{5.23}$$

その後，正規分布表から $\Phi(z_1)$ と $\Phi(z_2)$ を拾って差を求めるわけだが，このままだと無視できない誤差が入ることがある．図 5.9 にあるように，両端の x_1, x_2 が現れる確率はヒストグラムの区間幅が 1 の柱の面積なのだが，区間 $[z_1, z_2]$ の面積を求めたとすると，両端の柱が半分ずつしか含まれなくなる．これが誤差の原因だ．

そこで誤差を補正するために面積を求める区間を修正する．つまり，

$$\begin{aligned} z_1 &= \frac{x_1 - \frac{1}{2} - \mu}{\sigma} \\ z_2 &= \frac{x_2 + \frac{1}{2} - \mu}{\sigma} \end{aligned} \tag{5.24}$$

としてやればよい近似になる．これが半整数補正の内容である．

それでは，半整数補正をやらなくてもよい近似が成立するというのは，どういう条件の下でだろうか．それは式 (5.24) において 1/2 という補正によって z_1, z_2 が影響をほとんど受けないという条件だ．ということは，σ **が** 1/2 **よりもずっと大きい場合**ということになる．

たとえば $n = 400$ で $p = 0.5$ の場合，$\sigma = 100$ である．このときの 1/2 の補正の寄与は 0.5% しかないので，正規分布表を引いた時には意味がなくなる．

【章末問題】

問題 5–1 第一章で扱った 100 人の男子高校生の体重の平均と標準偏差から，平均値の前後の何 kg の範囲をとれば，半分の人数がそこに含まれることになるか計算して予測せよ．さらにその結果を実際のデータと比較してみよ．

問題 5–2 入試や模擬試験の個人成績を表すのによく使われる**偏差値 (standard score)** は，次のようにして算出される．全体の平均点を μ, 標準偏差を σ とした場合，点数が μ に等しいものを偏差値 50 とする．そして得点がそれより $\sigma, 2\sigma, \cdots$ だけ上回る点数を偏差値 $50, 60, \cdots$ とし，下回るほうについても同様に決める．

今，参加者が 12000 人の模擬試験で偏差値 58 を得た受験生がいたとする．この人の成績は全体で何番ぐらいに位置するかを求めなさい．

問題 5–3 日本人で A 型の血液をもつ人は 40% いる．24 人の日本人を集めた時に，A 型の血液を持つ人が 9 人以上 12 人以下である確率を求めよ．

問題 5–4 生まれる赤ちゃんが女の子である確率は 1/2 であるとする．ある年に 4 万人の赤ちゃんが生まれたとして，その男女比が平均からずれて，一方の性の比率が全体の 49% 以下または 51% 以上になる確率を求めなさい．

問題 5–5

1. 内閣支持率を無作為抽出によって調査したい． 真の支持率が 45.0 % であったとして，調査人数を 400 人とした場合に，調査結果として 得られる支持率が 45.0±2.5% の範囲になる確率を求めよ．

2. 上と状態における調査を，調査人数を 2000 人に増やして行いたい．このときに調査結果として得られる支持率が 45.0 ± 2.5 % の範囲になる確率を求めよ．

5.4 中心極限定理

前節で，二項分布の近似として正規分布が使えることがわかったが，それだけにとどまらず，もっと一般的な分布に対しても正規分布が成立することがしばしばある．それは次のような**中心極限定理 (central limit theorem)** によって保証される．

> 確率変数 $X_1, X_2, ..., X_n$ が互いに独立で，平均 μ, 分散 σ^2 をもつ分布に従っているとする．この時，平均
>
> $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5.25}$$
>
> をとり，
>
> $$Z = \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu) \tag{5.26}$$
>
> とすると，$n \to \infty$ の極限で Z は標準正規分布 $N[0,1]$ に従う．あるいは，式 (5.20) と見比べると，$\overline{X}$ は平均が μ, 分散が σ^2/n となるような正規分布 $N[\mu, \sigma^2/n]$ に従っていると言ってもよい．

この定理では，下線部の条件を満たす確率分布であれば，**もとの分布が正規分布でなくても**，そこからとった独立な多数の変数の平均が正規分布に従うことを保障している．これは驚くべき事実であり，正規分布のような「山なり」の分布でなくても，そこから何個かの確率変数をとって平均したものを集めると，正規分布になっているというのである．たとえば「平らな」分布である一様分布からも，そのようにして正規分布を作ることができることを，次の例で示した．

なおこの定理によって Z が正規分布とみなされるための n の値は，もとの分布の性質にもよるが，10 程度でもかなりよい近似を与える．多くの統計データの処理において正規分布を使った解析が威力を発揮するのは，この中心極限定理によって，確率的に振舞うデータから正規分布が実現することが保証されているからである[*8]．

問題 5–6 5.1.6 節 (68 ページ) に出てきた一様乱数は，分散が 1/12 である (式 (5.16))．一方，2 つの独立な確率変数の分散は，それぞれの分散を足し合わせることで求めること

[*8] このことについて考察しておこう．たとえば日本人の成人男子の身長のデータを正規分布しているとして扱っても，かなりよい近似になる．これは人の身長が，多くの遺伝子が独立に，あるいは共同して関与した結果であるとともに，胎内での発生過程に関わる母親の栄養状態や心理状況，生後の生育環境を構成する食物や気候や運動など，無数の要因が関わったものであり，それらの効果の総和として身長が決まることによる．つまり，独立した多数の要因によって決まる量は中心極限定理によって正規分布するのであるから，身長も正規分布する傾向をもつことになるのは自然なことである．もっとも，体重のデータは身長のおよそ 2 乗から 3 乗に比例するために，身長の場合よりも非対称な分布になる傾向が強くなる．

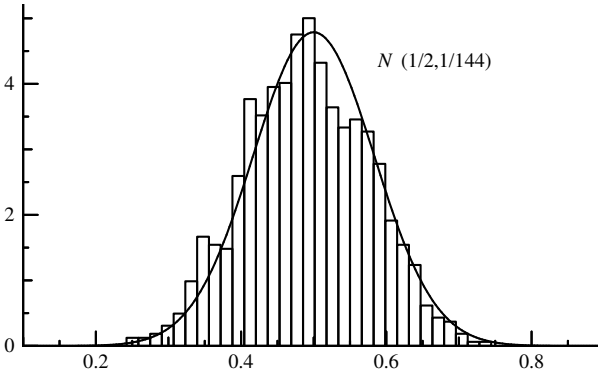

図 5.10　区間 $[0,1]$ の連続一様分布から 12 個の確率変数を取って作った平均がどのように分布するかを調べたヒストグラムと，中心極限定理で予想される正規分布の曲線.

ができる (式 (3.14)).

このことを利用すると，区間 $[0,1]$ の一様連続分布に従う乱数を 12 個足し合わせて，その和から 6 を引いた値はかなり標準正規分布に近い分布になる．そのことを説明しなさい.