第 4 章

二項分布

二項分布は，さまざまの状況で非常にしばしば現れる離散型確率分布である．また，二項分布を連続分布の極限へと延長すると，統計学において中心的な役割を果たす正規分布に導かれるし，「さみだれ式」の分布であるポアソン分布にもつながっている．

4.1 二項分布

4.1.1 二項分布を適用できるケース

血液型が A 型である人の割合は，日本人の場合，ほぼ 40% である．では，無作為に選んだ 10 人のうち 4 人が A 型である確率はどうなるだろうか．

10 人を次々につれてきて，最初から 4 人目までが A 型で，残りの 6 人がそれ以外である確率は，

$$0.4^4 \times (1-0.4)^6$$

となる．他の順序であっても，人数構成が同じであれば，この確率は変わらない．たとえば，最初に A 型が 2 人，次にそれ以外が 6 人，最後に A 型が 2 人でも，上の積の取り方の順序が変わって，

$$0.4^2 \times (1-0.4)^6 \times 0.4^2$$

となるだけで，確率は同じである．人を並べる順序が異なるということは互いに排反であるから，これらの確率は足し算されることになる．すなわち，A 型 4 人とそれ以外の 6 人を並べるやり方が何通りあるかを考えて，それを上の積に掛けてやれば，求めるべき確率が得られることになる．

ここで，A 型の人 4 人とそうでない人 6 人を並べるやり方は，

$$ {}_{10}C_4 = \frac{10!}{4!\ 6!} = 210 \ \text{通り}$$

である．従って，問題の答えは

$$ {}_{10}C_4 \times 0.4^4 \times (1-0.4)^6 = 210 \times 0.4^4(1-0.4)^6 = 0.251 $$

となる．

この導出にならって，n 回の試行において，確率 p であるような事象が x 回起きる確率関数は，一般に次の式で表される．

$$ f(x) = {}_nC_x p^x (1-p)^{n-x}, \quad (x = 0, 1, ..., n) \tag{4.1} $$

このような確率分布を**二項分布 (binominal distribution)** と呼び，しばしば $B[n, p]$ と記号で表される[*1]．二項分布は現実の現象においてしばしば登場する分布である．

例題 4–1　5 つの解答から正答 1 つを選択する方式の問題が 5 問出題されている．まったくランダムに答えを選択していった場合に，60% 以上の得点が得られる確率はどれほどになるかを計算せよ．

ひとつの問題で正答する確率は $1/5 = 0.2$ である．すると二項分布から，5 問中で正答が 5，4，3 である確率はそれぞれ，

$$ \frac{5!}{5!\ 0!}\ 0.2^5, \quad \frac{5!}{4!\ 1!}\ 0.2^4 \times 0.8, \quad \frac{5!}{3!\ 2!}\ 0.2^3 \times 0.8^2, $$

となるから，これらの和をとって，求める確率は 0.058 となる[*2]．

4.1.2　二項分布の期待値と分散

二項分布 $B[n, p]$ について，期待値と分散は次のようになる．平均については，意味を考えれば，この結果は自明である[*3]．なお，これらの導出は付録に与えてある．

$$ \mu = np \tag{4.2} $$

$$ \sigma^2 = np(1-p) \tag{4.3} $$

二項分布の期待値と分散は，後に出てくる正規分布との関係で非常に重要である．

---

*1　$B[n, p]$ という記号の中には，変数であるはずの x は含まれておらず，分布を特徴付ける n と p というパラメータだけが書かれていることに注意しておくこと．

*2 このケースでは，運よく合格できる確率が 5% は存在することになる．ここでもし問題の数がもっと多くて，たとえば 10 問だったとすると，合格の確率はどうなるだろうか．

*3　出現確率 p である現象が n 回の試行の中で平均して何回現れるかということであるから，当然 np という結果になる．もちろん，厳密な数学的導出によっても同じ結果が得られる．

4.1.3　二項分布の形

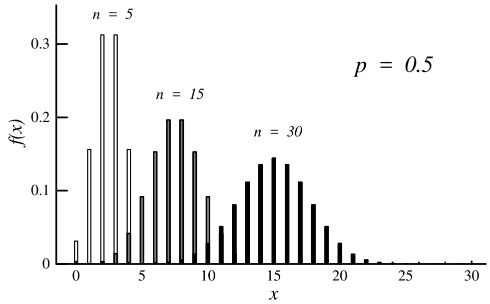

図 4.1　$p = 0.5$ の時の二項分布のようす

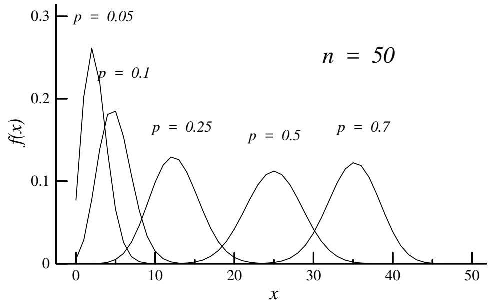

図 4.2　p を変化させた時の二項分布のようす．グラフの重なりを避けるために折れ線グラフで示した．

二項分布が実際にどのような分布になるかを，いくつかのグラフで示しておこう．図 4.1 は，$p = 1/2$ で n を 5, 15, 30 とした時の結果を図示したものである．いずれも中心に山を持ち，左右対称な分布になっている．このグラフは，コインを n 回投げる試行を行っては表の出る率を調べて，それを何度も繰り返してプロットしていった時に得られるものと同じものになるはずである．

図 4.2 は，p を変化させながら 50 回の試行を繰り返した時の二項分布のようすをプロットしたものである．np のところにモードを持つ分布になっていることが分かる．

4.2　多項分布

日本人の血液型分布は，A, O, B, AB 型の人の比率が，およそ 40%, 30%, 20%, 10% となっている．ここで 10 人の日本人がいたとすると，その血液型の構成が 2 人，4 人，2 人，2 人となっている確率はどれほどであるか．— **多項分布 (polynomial distribution)**

は，このようなケースに対して有効である．

この問題は二項分布の説明で述べたのと全く同様に考えればよい．まず，最初の 2 人が A 型，次の 4 人が O 型，続いて 2 人が B 型，最後の 2 人が AB 型となる確率は，

$$0.4^2 \times 0.3^4 \times 0.2^2 \times 0.1^2$$

で与えられる．しかし，この人数構成比になるための人の並び方の順序は，

$$\frac{10!}{2!4!2!2!}$$

通りであるから，これらの積を取れば，求める確率が得られる．

つまり，多項分布というのは，二項分布を多数の事象の場合について拡張したものに他ならない．この話を一般化すると，次のようになる．

排反な事象の完全な組，$E_1, E_2, ..., E_k$ があり, それぞれの実現確率が $p_1, p_2, ..., p_k$, ただし $p_1 + p_2 + ... + p_k = 1$ であるとする．n 回の試行において，それらが $x_1, x_2, ..., x_k$ 回ずつ実現する確率 $f(x_1, x_2, ..., x_k)$ は，次の式で与えられる．

$$f(x_1, x_2, ..., x_k) = \frac{n!}{x_1!\ x_2!\ \ldots x_k!} \times p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \tag{4.4}$$

4.3　ポアソン分布

4.3.1　二項分布からポアソン分布へ

二項分布から導かれるもうひとつの重要な分布は**ポアソン分布 (Poisson distribution)** である．これは，ひとつひとつの確率は低いものの，非常に多数の回数の試行が行われるような現象に対して適用される確率分布である．

例えば，ある窓口に対して，8 時間で平均 24 本の電話がかかってくるようなケースを考えよう．分刻みで見れば，480 分に 24 回の割合でかかってくるわけだから，ある 1 分の刻みの中でかかってくる確率は 0.05 であり，電話が来るかどうかを監視する回数は 480 回というわけである．このようなケースで，たとえば連続した 60 分間に電話が来る確率を計算するとすれば，$B[60, 0.05]$，すなわち $n = 60, p = 0.05$ であるような二項分布を使ってもよい．

しかし実際上は，${}_{60}C_x$ といった因子を計算することは困難であるから，数学的に扱いやすい形でこの問題を扱いたい．そこで登場するのがポアソン分布なのである．

4.3.2　ポアソン分布の導出

上に述べたように，二項分布に対して $n \to \infty, p \to 0,$ という極限操作を行う．すると，

$$f(x) = \lim_{n\to\infty, p\to 0} {}_nC_x p^x (1-p)^{n-x} = \frac{\mu^x}{x!} e^{-\mu} \tag{4.5}$$

となる．ただし $\mu = np$ としてある．これがポアソン分布である．この式の導出の詳細は付録に掲載した．ここに見るように，ポアソン分布は平均値 μ だけで関数の形が決まるので，$P[\mu]$ と表記されることもある．

上にも書いたように，ポアソン分布を使う理由は，大きな n について二項分布を計算することが困難であることによると考えてよい．実際，式 (4.5) の計算であれば，関数電卓でも求められるものである[*4]．

4.3.3　ポアソン分布の期待値と分散

▼期待値　ポアソン分布の期待値，つまり平均は μ そのものである．

▼分散　分散は，二項分布の場合の分散を表す式 (4.3) から，次のように簡単に導くことができる．

[*4] Windows に付いている電卓は関数電卓に切り替えて使えるので，この種の計算をこなすことができる．

$$\sigma^2 = \lim_{p\to 0} np(1-p) = np = \mu \tag{4.6}$$

つまり，ポアソン分布では期待値と分散はいずれも μ に等しいという面白い性質がある．

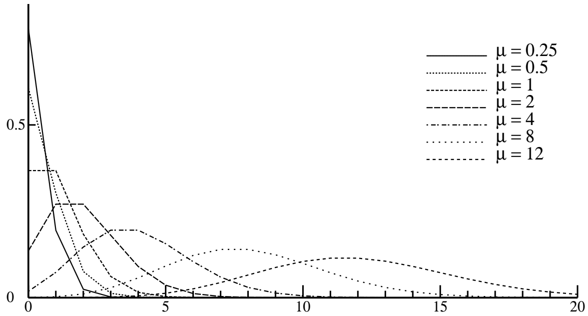

図 4.3　さまざまな μ の値に対するポアソン分布の形：ポアソン分布は離散分布なので確率は x が整数の時のみ値をもつ．しかし，その通りに描くと点だけのグラフになって見にくいので，点をつないでわかりやすくしている．

例題 4–2　60 分に平均して 2 回の電話が掛かってくる窓口がある．60 分の間に，この窓口に 5 本以上の電話が掛かってくる確率を求めよ．

ポアソン分布

$$f(x) = \frac{\mu^x}{x!}e^{-\mu}$$

で，$\mu = 2$ として考える．5 本以上の電話が来る事象は 4 本以内の電話が来るという事象の排反事象であるから，後者の確率すなわち $f(0)$ から $f(4)$ までの和を求めて，それを 1 から引けばよい．従って，

$$\begin{aligned} &1-(f(0)+f(1)+f(2)+f(3)+f(4)) \\ &\quad = 1 - e^{-2}\left(\frac{2^0}{0!}+\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!}+\frac{2^4}{4!}\right) \\ &\qquad = 0.053 \end{aligned}$$

となり，ほぼ 5% の確率となる．

4.3.4　ポアソン分布が適用される問題

上に述べたように，ポアソン分布は二項分布の n が大きく，かつ p が小さくて，平均値 np が有限であるようなケースに適用される．通常はおよそ $n > 50$ であることと，$n \approx 50$ に対して $np \leq 5$ 程度であれば適用可能であるとされている．

ここでは，ポアソン分布のもうひとつの特徴について言及しておこう．ポアソン分布の式，

$$f(x) = \frac{\mu^x}{x!} e^{-\mu}$$

を見ると，ここには n も p も姿を見せておらず，平均値 μ だけしかパラメータとして使われていない．ということは，ポアソン分布においては，多数回の試行や個々の事象は隠されてしまっており，ある事象の発生する平均的な頻度だけが表に出ているのである．このことはこの分布の適用される現象について，広範な一般性を与えることになり，有用な手法を提供することを意味する．

たとえば，夜空で流星が見られる回数について考えてみよう．流星群などが来ている特別な状況を除いて，長期間の観測を行えば，ある平均的な頻度が得られることになる．仮にそれを 1 時間当たり 5 回としよう．

私たちは流星が落ちる原因については何も知らないが，しかし地球の周りの宇宙空間でさまざまな偶然が重なった結果として，たまたま宇宙の塵が大気に突入することになったのだろうという程度の，無難でさして意味のない推定ぐらいはできるであろう．流星を産み出すことに関わる現象は，天空のどこかで時々刻々起きているであろうし，その中で，非常に「運の悪い」塵が叩き落されるとすると，これはポアソン分布を使ってよいケースになる．

流星の発生する仕組みについては全く無知である以上，その仕組みと関係して決まるはずの n や p に関する知識を得ることはできない．せいぜい，n は「時々刻々」の現象ということで非常に大きな数だろうとか，「運が悪いのが落ちる」のだろうから p は小さいだろうということを仮定してもかまわないだろうという程度の知識しかないのである．しかし，観測によって平均値 $\mu = np$ は確実に分かるから，それさえあればポアソン分布の式を利用することが可能になるというところに，この問題の面白さがある．

実際，$\mu = 5$ という値を使えば，たとえば 1 時間に 10 個以上の流星が観測できる確率を知ることができ，式から計算してみると，その確率は 0.03 である．そこでもし，あるときに 1 時間に 10 個の流星が観測されたとすると，それはきわめて稀な現象が起きているというよりは，週に 1,2 度くらいはそんなこともあるだろうという判断ができることになる．

このように，ポアソン分布は，現象の原因やメカニズムがわからなくても，さまざまの自然現象や社会現象の解析において，観測されたデータから確率的な推論を下すための，有力な武器を提供している．

【章末問題】

問題 4–1　ある大学では学生の 2/3 が関西出身である．この大学でランダムに 8 人を選んだとき，関西出身の学生が 4 人未満である確率を求めなさい．

問題 4–2　6 人がじゃんけんして，ぐう，ちょき，ぱあがそれぞれ 2 人ずつになる確率はどれだけか．

問題 4–3　ある中学校のクラスでは，1 週間に平均して 2.5 名の病気欠席者がいる．ある週とその次の週が連続して無欠席になる確率を求めなさい．