12.　二項分布

1.　反復試行の確率

●　反復試行の確率

1 回の試行で，事象 A の起こる確率が p であるとする。この試行の n 回の反復試行において，事象 A がちょうど x 回起こる確率を $p(x)$ とすると

$$p(x) = {}_nC_x\, p^x q^{n-x} \quad (0 \le x \le n\ ;\ q = 1-p)$$

（注）必然的に　$p(0) + p(1) + \cdots + p(n) = 1$　となる。

(1)　上記は，高校数学 A で学習する。証明を確認する。1 回の試行で，A が起こる確率は p，A が起こらない確率は $1-p$ である。$q = 1-p$ とおくと，A の余事象 $\overline{A}$ の確率が q である。

n 回の反復試行において，各回の結果を〇で表すと，A がちょうど x 回起こる場合は，次の形である。

（※）　〇 〇 ……　〇　（x 個の〇は A，残り $n-x$ 個の〇は $\overline{A}$）

まず，この場合の数を求める。これは，n 個の〇から，A となる x 個の〇の選び方の数であるから，${}_nC_x$ 通りである。この ${}_nC_x$ 通りのうち，どの場合の確率も，（※）の形から，

（x 個の p の積）×（$n-x$ 個の q の積）$= p^x q^{n-x}$

従って，求める確率は，${}_nC_x\, p^x q^{n-x}$　となる。

(2)　次が成立する。

$$p(0) + p(1) + \cdots + p(n) = 1$$

なぜなら，n 回の反復試行を行うとき，x が 0〜n のいずれかの整数になる確率は，上式の左辺である。一方，x は必ず 0〜n のいずれかの整数になるので，その確率は 1 である。

あるいは，次の二項定理からでも確認できる。

$$1 = (p+q)^n = \sum_{x=0}^{n} {}_nC_x\, p^x q^{n-x} = \sum_{x=0}^{n} p(x)$$

(3)　「反復試行の確率」は，昔の高校数学では「独立試行の定理」と呼んだ。

試行 T を n 回繰り返したとき，毎回の試行結果が他の回の試行結果に影響を及ぼさないとき，「T を n 回続けて行う」という試行を，「反復試行」または「n 回の反復試行」という。つまり，試行を繰り返したとき，各回の試行が互いに独立の場合，この試行の繰り返しを反復試行というわけである。サイコロを n 回投げる，コインを n 回投げるなどは，n 回の反復試行である。

(4)　母集団に対する「大きさ n の復元無作為抽出」は，n 回の反復試行である。母集団から無作為に 1 個とるという試行を n 回繰り返すが，復元抽出なので，各回の試行は互いに独立で

ある。

(5)　1個のサイコロをn回投げる試行と，n個のサイコロを同時に1回投げる試行は，同じ試行である。よって，複数の同じ試行を同時に行い，それぞれの試行が互いに独立であれば，反復試行の確率が適用できる。

■ 例題１

反復試行の確率が理解できないときは，簡単な具体例で理解すればよい。

例えば，1個のサイコロを3回投げる反復試行において，1の目が出る回数をXとする。

このとき，$X=2$となるのは，3回のうち2回だけ1の目が出る場合であり，その場合の数は 下表のように ${}_3C_2=3$（通り）である。

1回目	2回目	3回目	確率
○	○	×	$\frac{1}{6}\times\frac{1}{6}\times\frac{5}{6}=\left(\frac{1}{6}\right)^2\frac{5}{6}$
○	×	○	$\frac{1}{6}\times\frac{5}{6}\times\frac{1}{6}=\left(\frac{1}{6}\right)^2\frac{5}{6}$
×	○	○	$\frac{5}{6}\times\frac{1}{6}\times\frac{1}{6}=\left(\frac{1}{6}\right)^2\frac{5}{6}$

○ 1の目　　× 1以外の目

この 3 通りのどの場合においても，その事象が起こる確率は $\left(\frac{1}{6}\right)^2\frac{5}{6}$ であるから，$X=2$となる確率は

$$p(2)=P(X=2)={}_3C_2\left(\frac{1}{6}\right)^2\frac{5}{6}$$

である。同様に考えると

$$p(x)=P(X=x)={}_3C_x\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{3-x}\quad (x=0,1,2,3)$$

■ 例題２

1個のサイコロを10回投げたとき，1または2の目がちょうど4回出る確率を求めよ。

＜解＞

1個のサイコロを1回投げたとき，1または2の目が出る確率は$p=2/6=1/3$である。よって，$q=1-1/3=2/3$であるから，求める確率は次のとおり。

$$p(4)={}_{10}C_4\,p^4q^{10-4}={}_{10}C_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^{10-4}$$

2. 二項分布の定義

● **定義（二項分布）**

p, q は，$0 \leq p \leq 1$，$q = 1 - p$ であるような実数とする。

離散的確率変数 X のとる値が，$X = 0, 1, 2, \cdots, n$ であり，$X = x$ となる確率が

$$p(x) = P(X = x) = {}_nC_x\, p^x q^{n-x} \quad (x = 0, 1, \cdots, n)$$

であるとき，X は二項分布 $B(n, p)$ に従うといい，$X \sim B(n, p)$ で表す。

X の確率分布表は次のようになる。

X	0	1	$\cdots$	x	$\cdots$	n	計
P	${}_nC_0\, q^n$	${}_nC_1\, pq^{n-1}$	$\cdots$	${}_nC_x\, p^x q^{n-x}$	$\cdots$	${}_nC_n\, p^n$	1

● **二項分布の典型例**

(A) 1 回の試行で，事象 A の起こる確率が p であるとする。この試行の n 回の反復試行において，事象 A の起こる回数を X とすれば，$X \sim B(n, p)$ である。

(B) 母集団において，性質 E をもつ個体の個数の割合を p とする。大きさ n の標本の復元無作為抽出を行い，抽出した標本に含まれる性質 E をもつ個体の個数を X とすると，$X \sim B(n, p)$ である。

(1) 上記の二項分布の典型例は，二項分布の定義と考えてもよい。(A) は，反復試行の確率より自明である。

(2) (B) については，性質 E とは，男性，赤玉，不良品，喫煙者など，個体に関する特性のことである。例えば，母集団を機械の部品の集まりとし，不良品の割合を p とする。（部品が全部で 1000 個，不良品が 30 個あれば，不良品の割合は $p = 30/1000 = 0.03$ である。）母集団から無作為に 1 個の部品を選ぶという試行を行うと，選ばれた部品が不良品である確率は p である。この試行を復元で n 回行うと，n 回の反復試行になるので，不良品が選ばれる回数（＝大きさ n の標本に含まれる不良品の個数）X は，二項分布 $B(n, p)$ に従う。

(3) 二項分布は，離散的な確率分布において最も有名であり，重要な分布でもある。定義の $p(x)$ が X の確率関数になる。（分布の形はテキスト p.141 を参照）

(4) 二項分布は，何らかの現象を n 回観測したとき，特定の事象 A が何回起こるのかを示した分布であると理解すればよい。二項分布の応用範囲は広く，「製品の不良率，政権への支持率，テレビの視聴率」などの比率に関する統計的推測の際に用いられる。

■ **例 1（サイコロ投げ）**

1 個のサイコロをn回投げたとき，1 の目が出る回数を X とする。あるいは，n個のサイコロを同時に投げたとき，1 の目が出るサイコロの個数を X とする。

$$X \sim B(n,\ \frac{1}{6}),\qquad P(X = x) = {}_nC_x\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{n-x}$$

■ 例 2（コイン投げ）

1 枚のコインをn回投げたとき，表が出る回数を X とする。あるいは，n枚のコインを同時に投げたとき，表が出るコインの枚数を X とする。

$$X \sim B(n,\ \frac{1}{2}),\qquad P(X = x) = {}_nC_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x} = {}_nC_x\left(\frac{1}{2}\right)^n$$

■ 例 3（観測回数）

1 回の観測でカラスを発見する確率が p であるとき，n回観測したとき，カラスを発見する回数を X とする。各回の観測結果が他の回の観測結果に影響を及ぼさないと判断できれば，n回の反復試行と考えてよいので，$X \sim B(n,\ p)$ となる。

■ 例 4（復元無作為抽出）

赤玉 3 個，白玉 4 個が入っている袋から 1 個の玉を無作為に取り出し，色を調べてから袋に戻す。これを 10 回繰り返すとき，赤玉を取り出す回数を X とする。これは，大きさ 10 の標本の復元無作為抽出である。1 個の赤玉を取り出す確率は，毎回 3／7 であるので，$X \sim B(10,\ 3/7)$ である。

■ 例 5（不良品の個数）

不良率が 3%である部品の山から，10 個の部品を無作為に抽出したとき，この 10 個に含まれる不良品の個数を X とすると，$X \sim B(10,\ 0.0.3)$ である。

（注意）

10 個の部品の抽出は，1 個ずつとると考える。すなわち，「無作為に 1 個とる」という試行を 10 回繰り返すと考える。

その場合，復元か非復元が問題になるが，現実の抽出は非復元である。しかし，この種の問題では，部品はたくさんあると考え，部品の山は無限母集団と見なしていく。そうすれば，非復元であっても，大きさ 10 の復元無作為抽出と考えてよい。このように考えれば，10 回の反復試行になり，毎回不良品を選ぶ確率は 0.03 になるので，$X \sim B(10,\ 0.0.3)$ となる。

■ 例 6（割合）

ある集団には，左ききの人が 10%いるといわれている。この集団から 20 人を選んで左ききかどうかを調べたとき，20 人の中の左ききの人の人数を X とすると，$X \sim B(20,\ 0.1)$ である。

（注意）

集団の人数も不明，選び方も不明であるので，実際の判断は慎重になる必要があるが，統計

学の演習問題としては前の例と同様に考えて， $X \sim B(20,\ 0.1)$ と判断する。

■　例 7（死亡数）

ある薬品をマウスに対して一定量注射するとき，死亡率が 1／5 であるとすれば，無作為に 10 匹抽出して実験するとき，その死亡数を X とすると， $X \sim B(10, 1/5)$ である。

（注意）

マウスの全体は，無限母集団と見なす。死亡率が 1／5 とは，その母集団の中で，注射をすれば死ぬマウスが 20%（＝1／5），死なないマウスは 80%いるという意味である。

■　例 8（出生率）

日本では，男児の出生率は約 0.51 である。子どもが 5 人いる世帯において，男児の人数を X とすると， $X \sim B(5, 0.51)$ である。

（注意）

毎年の男児の出生率（＝生まれた男児の数／生まれた子どもの数）は，ほぼ 0.51 である。よって，子どもの全体の 51%は男児と見なし， $X \sim B(5, 0.51)$ と考えることができる。

3.　二項分布の平均・分散・標準偏差

●　定理（二項分布の平均・分散・標準偏差）

$X \sim B(n,\ p)$ のとき，次が成り立つ。

①　X の平均　$E(X) = np$

②　X の分散　$V(X) = npq$（$q = 1 - p$）

③　X の標準偏差　$\sigma(X) = \sqrt{V(X)} = \sqrt{npq}$

上記は，定義に従って証明するのは面倒である。例えば， X の平均 $E(X)$ は， X の確率分布表から

$$E(X) = \sum_{x=0}^{n} x\,{}_nC_x\,p^x q^{n-x}$$

となるが，この右辺を式変形するのは面倒である。

実は，上記は標本変量を導入すれば，すでに証明されている（後述）。ここでは，高校数学 B の教科書に書かれている証明（よくやる方法）を記しておく。

サイコロ投げを考えれば十分である。1 個のサイコロを n 回投げて，1 の目が出る回数を X とすれば，

$$X \sim B(n,\ p),\quad p = 1/6$$

次に，1 個のサイコロを n 回投げるという試行において，

i 回目に 1 の目が出たときは　　$X_i = 1$

i 回目に 1 の目が出ないときは　$X_i = 0$

とおき， n 個の確率変数 X_1，X_2，…，X_n を定める。

従って, これらの変数の値は 0 または 1 であり, $X_2 = 0$ であれば 2 回目に 1 の目は出ない, $X_3 = 1$ であれれば，3 回目に 1 の目は出たという意味になる。

よって，和 $X_1 + X_2 + \cdots + X_n$ の値は，n 回投げたときの 1 の目が出る回数を表すことになり，

$$X = X_1 + X_2 + \cdots + X_n$$

となる。

一方，X_1，X_2，…，X_n が独立であることは自明であり，

どの X_i についても，その確率分布は右のようになり，

$$P(X_i = 1) = p$$
$$P(X_i = 0) = q \quad (q = 1 - p)$$

X_i	0	1	計
P	q	p	1

よって，

$$E(X_i) = 0 \times q + 1 \times p = p$$

$$V(X_i) = E(X_i^2) - E(X_i)^2 = 0^2 \times q + 1^2 \times p - p^2$$
$$= p(1 - p) = pq$$

従って，

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$
$$= p + p + \cdots + p = np$$

$$V(X) = V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$$
$$= pq + pq + \cdots + pq = npq$$

■ 例題

(1) 確率変数 X が二項分布 $B(5, 1/6)$ に従うとき，X の平均，分散，標準偏差を求めよ。

(2) 赤玉 3 個と黒玉 1 個が入っている袋から 1 個の玉を無作為に取り出し, 色を調べてから袋に戻す。これを 100 回繰り返すとき，赤玉を取り出す回数を X とする。X の平均，分散，標準偏差を求めよ。

＜解＞

(1) $X \sim B(5, \frac{1}{6})$ であるから，

X の平均は $E(X) = 5 \times \frac{1}{6} = \frac{5}{6}$

X の分散は $V(X) = 5 \times \frac{1}{6} \times \frac{5}{6} = \frac{25}{36}$

X の標準偏差は $\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{25}{36}} = \frac{5}{6}$

(2) 袋から 1 個の玉を取り出したとき，それが赤玉である確率は 3／4 であるから，

$X \sim B(100,\ \frac{3}{4})$ である。従って，

X の平均は $E(X) = 100 \times \frac{3}{4} = 75$

X の分散は $V(X) = 100 \times \frac{3}{4} \times \frac{1}{4}$

X の標準偏差は $\sigma(X) = \sqrt{V(X)} = \sqrt{100 \times \frac{3}{4} \times \frac{1}{4}} = \frac{5\sqrt{3}}{2}$

4. 二項分布は正規分布に近づく

● **定理**

n を大きくすると，二項分布 $B(n,\ p)$ は 正規分布 $N(np,\ npq)$（$q = 1 - p$）に近づく。

＜説明＞

サイコロ投げの回数 n を増やしていくと，二項分布 $B(n,\ p)$ は 正規分布 $N(np,\ npq)$ に近づくことが，中心極限定理より簡単に確認できる。上記は，非常に重要な定理である。

母集団

白玉 3 個	○ ○ ○
赤玉 1 個	△

具体例で確認してみよう。右の図のように，母集団は 4 個の玉からなり，白玉が 3 個，赤玉が 1 個あるとする。

この母集団から無作為に 1 個の玉を取り出したとき，それが白玉であるという事象を A とすれば，当然

$$P(A) = 3/4$$

従って，無作為に 1 個とるという試行を復元で n 回繰り返したとき，n 回の反復試行になるので，n 回のうち白玉を取り出す回数を Y とすれば，

$$Y \sim B(n,\ p),\quad p = 3/4$$

このことは，無作為標本でも考えることができる。

まず，母集団の個体の特性値 X を次のように定める。

$$X = \begin{cases} 1 & （白玉の場合） \\ 0 & （赤玉の場合） \end{cases}$$

X	0	1	計
P	q	p	1

このとき，母集団分布（X の確率分布）は，右のようになる。

ここで，$p = 3/4$，$q = 1 - p = 1/4$ である。

また，容易に次がわかる。

母平均 $E(X) = p$，　母分散 $V(X) = pq$

次に，この母集団に対して，

試行T：大きさnの標本の復元無作為抽出

を考え，大きさnの標本変量（X_1，X_2，…，X_n）を考える。

以下のことが容易にわかるだろう。

試行Tは，母集団から無作為に1個とるという試行を復元でn回繰り返すことなので，n回の反復試行になる。X_1，X_2，…，X_nの値は0または1であり，例えば，$X_2=1$とは，2回目には白玉を取り出したという意味になる。

従って，和 $X_1+X_2+\cdots+X_n$ の値は，次のことを意味する。

大きさnの無作為標本に含まれる白玉の個数
（n回の反復試行において白玉を取り出した回数）

よって，

$$Y=X_1+X_2+\cdots+X_n$$

となる。

一方，標本平均$\overline{X}$の定義から，$\overline{X}=\dfrac{1}{n}Y$ であり，p.80 より，

$$E(\overline{X})=E(X)=p,\quad V(\overline{X})=\frac{V(X)}{n}=\frac{pq}{n}$$

このことからも，次の計算により，$E(Y)=np$，$V(Y)=npq$ が証明される。

$$E(Y)=E(n\overline{X})=nE(\overline{X})=np$$
$$V(Y)=V(n\overline{X})=n^2V(\overline{X})=n^2\cdot\frac{pq}{n}=npq$$

さて，中心極限定理より，nを大きくすると，標本平均$\overline{X}$の分布は，正規分布

$$N(E(\overline{X}),V(\overline{X}))\quad すなわち，\quad N(p,\frac{pq}{n})$$

に近づいていく。

よって，$\overline{X}$の1次変換である $Y=n\overline{X}$ も，正規分布

$$N(E(Y),V(Y))\quad すなわち，\quad N(np,npq)$$

に近づいていく。

5.　二項分布の正規分布による近似

● **定理（ド・モアブルーラプラスの定理）**

(1)　nが十分大きければ，二項分布$B(n,\ p)$は，近似的に正規分布

$$N(np,\ npq)\quad (q = 1 - p)$$

に従うと考えてよい。（注）十分大きなnの基準例：$n \geq 100$

(2)　従って，$X \sim B(n,\ p)$（$n \geq 100$）のときは，

$$X \sim N(np,\ npq)$$

である。このとき，Xを標準化すれば，

$$Z = \frac{X - np}{\sqrt{npq}} \sim N(0,\ 1)$$

(3)　**（正規近似の半整数補正）**

二項分布$B(n,\ p)$の確率計算を，正規分布$N(np,\ npq)$で近似して計算することを，正規近似で計算するという。その計算では，半整数補正を行い，以下の右辺のように計算すると精度が高まる。a，bは整数である。

① $P(X \geq a) = P(X > a - 0.5)$

② $P(X \leq b) = P(X < b + 0.5)$

③ $P(a \leq X \leq b) = P(a - 0.5 < X < b + 0.5)$

(1)　$X \sim B(n,\ p)$ のとき，Xの平均はnp，分散はnpqである。nが十分大きければ，Xの分布は正規分布と見なしてよいので，$X \sim N(○,\ △)$となる。このとき，○はXの平均np，△はXの分散npqである。

(2)　二項分布$B(n,\ p)$の確率計算は，nが大きくなると，手計算では無理である。例えば，$X \sim B(50, 0.3)$ のとき，$P(X = 20) = {}_{50}C_{20}(0.3)^{20}(0.7)^{30}$ の計算は，このままではできない。ところが，nが十分大きいときは，$B(n,\ p)$は正規分布$N(np,\ npq)$に近くなる。よって，二項分布の確率計算を，正規分布で近似して計算できる。この計算を，正規近似で計算するという。

(3)　「nが十分大きいとき」とあるが，nに関する統一的な基準はない。以下のような基準はあるが，計算結果の精度をどこまで求めるかで，基準が異なってくる。通常は，$n \geq 100$であれば，近似計算はOKと思ってよい。

（基準A）：$n \geq 100$

（基準B）：$n > 30$，$np > 5$，$nq > 5$

（基準C）：$np > 5$，$nq > 5$

(4)　二項分布 $B(n,\ p)$ に従う確率変数 X は，とる値は整数値である。よって，

$$P(20 \le X \le 30)$$

のような確率を計算することになるが，正規分布と見なして計算するときは，

$$P(20 - 0.5 < X < 30 + 0.5) \quad \text{すなわち} \quad P(19.5 < X < 30.5)$$

のように，前後に 0.5 ずつ範囲を広げて計算するとよい。これを，半整数補正という。

高校数学 B でも正規近似の計算は登場するが，半整数補正を行わずに計算している。離散分布は長方形の集まりで表現でき，連続分布は曲線なので，半整数補正を行うと，計算結果の精度が高まる。

■　例題

1 個のサイコロを 720 回投げるとき，次の確率を正規近似で求めよ。

(1)　1 の目の出る回数が 140 回以上になる確率

(2)　1 の目の出る回数が 115 回以上 125 回以下になる確率

＜解＞

1 の目の出る回数を X とすると， $X \sim B(720,\ 1/6)$ であるから，

$$E(X) = 720 \times \frac{1}{6} = 120,\quad V(X) = 720 \times \frac{1}{6} \times \frac{5}{6} = 100$$

$$\sigma(X) = \sqrt{100} = 10$$

ここで，試行回数 n は $n \ge 100$ を満たすので， X は正規分布で近似できる。すなわち，

$$X \sim N(E(X),\ V(X)) \quad \text{すなわち} \quad X \sim N(120,\ 100)$$

と見なしてよい。 X を標準化すると，

$$Z = \frac{X - E(X)}{\sigma(X)} = \frac{X - 120}{10} \sim N(0,\ 1)$$

以下，半整数補正を行って計算する。

(1)　$P(X \ge 140) = P(X > 140 - 0.5) = P(X > 139.5)$

$$= 1 - P(X \le 139.5) = 1 - P\left(Z \le \frac{139.5 - 120}{10}\right)$$

$$= 1 - P(Z \le 1.95) = 1 - 0.9744 = 0.0256$$ （答）

(2)　$P(115 \le X \le 125) = P(115 - 0.5 < X < 125 + 0.5)$

$$= P(114.5 < X < 125.5) = P\left(\frac{114.5 - 120}{10} < Z < \frac{125.5 - 120}{10}\right)$$

$$= P(-0.55 < Z < 0.55) = 2 \times P(0 < Z < 0.55)$$

$$= 2 \times (0.7088 - 0.5) = 2 \times 0.2088 = 0.4176$$ （答）