

2項分布, ポアソン分布, 幾何分布

今日から3回の講義で, 統計学でよく使われる確率分布モデルをいくつか説明します. 今日は, ここまでの講義で例としてあげた「階級値」や「当たり本数」のように, 確率変数の値がとびとびの値になる場合の確率分布モデル¹をいくつか紹介します. 本文中の式の導出は, 後につけた付録を見てください.

ベルヌーイ試行と2項分布

問題1: あるくじびきは, 5%の確率で当たりが出るとします. このくじを5本ひいたとき, その中の2本以上が当たりである確率はいくらですか.

この問題のように,

- 「当たり・はずれ」, 「成功・失敗」など, 2種類の結果のどちらかが必ず起きるような実験 (あるいは観測, 検査) があり,
- この実験を何度も繰り返すとき, 各実験での「成功」「失敗」それぞれの確率はつねに一定で,
- 1つの実験の結果が他の実験に影響を及ぼさない (各実験が独立),

となります. このような仮定ができるような実験を**ベルヌーイ試行**といいます. ベルヌーイ試行で起こりうる2種類の結果を「成功」, 「失敗」で表し, 成功する確率が p , 失敗する確率が $1-p$ とします. このとき, 「 n 回のベルヌーイ試行を行うとき, 成功する回数」を X とします.

上のベルヌーイ試行について, X のしたがる確率分布を考えてみましょう. ここで, 成功を○, 失敗を●の記号で表すことにします. こうすると, 「 n 回の試行を行うとき, ちょうど x 回成功する」という状態は, 「○が x 個, ●が $n-x$ 個」からなる○●の列で表されます. 成功する確率が p , 失敗する確率が $1-p$ で, それぞれの試行が独立ですから, 「○が x 個, ●が $n-x$ 個」という列の1つが起きる確率は $p^x(1-p)^{n-x}$ です. そして, そのような列が何種類あるかは「 n 個から x 個とる組み合わせの数」で表わされ, これは ${}_nC_x$ という記号で表されます.

よって, 「 n 回の試行を行うとき, x 回成功する確率」 $P(X=x)$ は, 次のように表されます.

$$P(X=x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad (1)$$

この確率を, 成功回数の例 x が最小の0回するときから最大の n 回するときまですべてについて求めると, 成功回数を表す確率変数 X がしたがる確率分布が求められます. この確率分布は, 「 n 回の試行中 x 回成功する確率を, ベルヌーイ試行の仮定から求めたもの」ですから, 確率分布モデルです. この確率分布モデルには**2項分布**という名前があり, $B(n, p)$ という記号で書くこともあります. ここで, n と p はパラメータです. つまり, 「 (n, p) 」と書いてあるのは, 試行回数 n と1回の試行での成功確率 p がわかれば, どの成功回数 x についても「 x 回成功する確率」は計算できる, という意味です.

¹離散型確率分布モデルといいます. 詳しくは, 次回の講義で「連続型確率分布モデル」と一緒に説明します.

ある確率変数 X が2項分布 $B(n, p)$ にしたがうとき、その期待値 $E(X)$ と分散 $V(X)$ は次のように表されます。

$$E(X) = np, V(X) = np(1 - p) \quad (2)$$

試行1回あたりの成功の確率が p で、それを n 回行うわけですから、「 n 回試行したとき成功する回数の期待値」が np であることは、直感的にわかります。また、 $p = 1/2$ の時分散が最大になることは、「成功する回数」の予測が、 $p = 1/2$ のとき、一番しにくいことを示しています。

ポアソン分布とポアソン過程

問題2：ある電話は、1分間の間にある着信の回数の期待値は3回であるという。この電話に1分間に5回以上の着信がある確率はいくらか。

この問題で、1分間を非常に細かく分けて、分割された各時間帯では「着信がないか、1回あるかどうか」であるとしましょう。つまり、1つの時間帯に着信が2回あることはない、というくらい細かく分けるわけです。こうすると、各時間帯では「着信がある→成功、ない→失敗」というベルヌーイ試行が起こっていると考えられます。

ここで、1分間を非常に細かく分けたので、試行回数 n は $n \rightarrow \infty$ となります。しかし、1分間の着信の回数の期待値は3回で一定です。前節で説明した通り、「分割された各時間帯で着信がある」確率を p とすると、1分間の着信の回数の期待値は np で、 $n \rightarrow \infty$ なのに np が一定なのですから、 p が $p \rightarrow 0$ になっています。

このように、前節の2項分布において $n \rightarrow \infty$ とし、さらに期待値 $E(X) = np = \lambda$ で一定値であるとし、このような状況での2項分布の極限を**ポアソン分布**といい、定数 λ をパラメータとして $Po(\lambda)$ で表します。また、この問題のようにポアソン分布が適用できる実験を**ポアソン過程**といいます。ポアソン分布の確率分布は、次のように表されます。

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (3)$$

また、期待値 $E(X)$ と分散 $V(X)$ は次のように表されます。

$$E(X) = \lambda, V(X) = \lambda \quad (4)$$

つまり、期待値も分散も同じになります。これは、2項分布の期待値と分散の式(2)で、 $np = \lambda, (1-p) \rightarrow 1$ の極限を考えれば理解できます。

幾何分布

問題3：ある町では、1年間に台風が襲われて被害が出る確率が0.04であるという。つぎにこの災害が起こるまでの年数の期待値はいくらか。また、その標準偏差はいくらか。さらに、この災害が10年以内に起こる確率はいくらか。

2項分布について学ぶついでに、2項分布と同じベルヌーイ試行について、別の値を確率変数とみた場合を考えてみましょう。上の問題で、ある年に台風が来るか来ないかというのはベルヌーイ試行になっています（この問題では、今年台風が来る確率は去年台風が来たかどうかには無関係であることに注意してください）。ここで、「何回も繰り返してベルヌーイ試行を行うとき、初めて成功するまでの回数」を確率変数 X とします。各試行の成功の確率が p であるとき、「最初の成功が起こるまで x 回かかる確率」 $P(X = x)$ は、「 $(x - 1)$ 回失敗して、その後 1 回成功する確率」ですから

$$P(X = x) = p(1 - p)^{x-1} \quad (5)$$

となります。(5) 式であらわされる確率分布を**幾何分布**といいます。幾何分布の期待値と分散は

$$E(X) = \frac{1}{p}, \quad V(X) = \frac{1-p}{p^2} \quad (6)$$

となります。例えば、1 回成功する確率が $1/5$ なら、はじめて成功するまでに平均 5 回かかる、というわけです。幾何分布は、 x を例えば 1 年、2 年、... のような「離散的な」時間としたとき、初めて成功するまでの待ち時間の確率分布ですから、**離散的待ち時間分布**とも呼ばれます²。

なお、幾何分布で「はじめて成功するまでの試行回数が x 回以内である確率」 $P(X \leq x)$ は

$$P(X \leq x) = 1 - (1 - p)^x \quad (7)$$

となります。これは、(5) 式が等比数列の形になっているため、等比数列の和の公式を使ってただちに求められます。

今日の演習

1. 本文の問題 1 に、上で説明した知識を使って答えるために必要な仮定を述べたうえで、答えてください。
2. 本文の問題 2 で、「1 分間にかかってくる電話の回数」を確率変数 X として X の確率分布を記述し、問題に答えてください。ただし $e = 2.718$ とします。
3. 本文の問題 3 に答えてください。また、「25 年に 1 回の大災害」は 25 年先にしか来ないといえるでしょうか？

付録

2項分布の期待値と分散の導出

2項分布の頻度関数からモーメント母関数を求めると

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{x=0}^n e^{tx} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \frac{n!}{x!(n-x)!} (pe^t)^x (1-p)^{n-x} \end{aligned} \quad (A1)$$

² 「連続型待ち時間分布」というものもあります。第 7 回の講義で説明します。

となります。 (A1) 式の下の式は2項展開そのものですから、

$$M_X(t) = (pe^t + (1-p))^n \quad (A2)$$

です。これを t で1回および2回微分すると

$$\begin{aligned} M'_X(t) &= npe^t (pe^t + (1-p))^{n-1} \\ M''_X(t) &= np \left[e^t(n-1)(pe^t + (1-p))^{n-2} pe^t + e^t (pe^t + (1-p))^{n-1} \right] \\ &= npe^t \left[\{pe^t(n-1) + (pe^t + (1-p))\} (pe^t + (1-p))^{n-2} \right] \\ &= npe^t \left[(npe^t + (1-p)) (pe^t + (1-p))^{n-2} \right] \end{aligned} \quad (A3)$$

となります。よって $E(X) = M'_X(0) = np$ です。また、 $E(X^2) = M''_X(0) = np((1-p) + np)$ となるので、 $V(X) = E(X^2) - \{E(X)\}^2 = np((1-p) + np) - (np)^2 = np(1-p)$ となります。なお、ここでは

$$\begin{aligned} V(X) &= \sum_x (x - E(X))^2 P(X = x) \\ &= \sum_x (x^2 - 2xE(x) + \{E(X)\}^2) P(X = x) \\ &= \sum_x x^2 P(X = x) - 2E(X) \sum_x x P(X = x) + \{E(X)\}^2 \sum_x P(X = x) \\ &= E(X^2) - 2E(X) \cdot E(X) + \{E(X)\}^2 \cdot 1 = E(X^2) - \{E(X)\}^2 \end{aligned} \quad (A4)$$

という関係を用いました。

2項分布の極限としてのポアソン分布の導出

2項分布の式 ((1) 式) を次のように変形します。

$$\begin{aligned} P(X = x) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{n(n-1)\cdots(n-(x-1))}{x!} p^x (1-p)^{n-x} \\ (\text{分母分子を } n^x \text{ 倍}) &= \frac{n(n-1)\cdots(n-(x-1))}{n^x x!} (np)^x (1-p)^{n-x} \\ (np = \lambda \text{ だから}) &= \frac{n(n-1)\cdots(n-(x-1))}{n \cdot n \cdots n} \cdot \frac{\lambda^x}{x!} (1-p)^{n-x} \\ &= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-(x-1)}{n} \cdot \frac{\lambda^x}{x!} (1-p)^{n-x} \\ &= \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \cdot \frac{\lambda^x}{x!} (1-p)^{n-x} \\ &= \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right)}{(1-p)^x} \cdot \frac{\lambda^x}{x!} (1-p)^n \end{aligned} \quad (A5)$$

ここで、 $z = -p$ とおくと

$$\begin{aligned} (1-p)^n &= \left[(1-p)^{-\frac{1}{p}} \right]^{-np} = \left[(1-p)^{-\frac{1}{p}} \right]^{-\lambda} \\ &= \left[(1+z)^{\frac{1}{z}} \right]^{-\lambda} \end{aligned} \quad (A6)$$

と変形でき、 $np = \lambda$ に固定して $n \rightarrow \infty$ とすると $p \rightarrow 0, z \rightarrow 0$ ですから、

$$\lim_{n \rightarrow \infty} (1-p)^n = \lim_{p \rightarrow 0} \left[(1-p)^{-\frac{1}{p}} \right]^{-np} = \lim_{z \rightarrow 0} \left[(1+z)^{\frac{1}{z}} \right]^{-\lambda} = e^{-\lambda} \quad (\text{A7})$$

です³。また、 $n \rightarrow \infty$ のとき $p \rightarrow 0$ であることより、

$$\lim_{n \rightarrow \infty} \frac{(1 - \frac{1}{n})(1 - \frac{2}{n}) \cdots (1 - \frac{x-1}{n})}{(1-p)^x} = 1 \quad (\text{A8})$$

です。(A5)式に(A7)(A8)式を適用すると

$$\lim_{n \rightarrow \infty} P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (\text{A9})$$

となり、二項分布の極限としてのポアソン分布が導かれます。

幾何分布の期待値と分散の導出

幾何分布の式 ((5) 式) から

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} = p \sum_{x=1}^{\infty} x(1-p)^{x-1} \\ &= p \left[1 \cdot (1-p)^0 + 2 \cdot (1-p)^1 + 3 \cdot (1-p)^2 + \cdots + k(1-p)^{k-1} + \cdots \right] \end{aligned} \quad (\text{A10})$$

となるので、

$$(1-p)E(X) = p \left[1 \cdot (1-p)^1 + 2 \cdot (1-p)^2 + 3 \cdot (1-p)^3 + \cdots + k(1-p)^k + \cdots \right] \quad (\text{A11})$$

となります。よって、

$$E(X) - (1-p)E(X) = p \left[(1-p)^0 + (1-p)^1 + (1-p)^2 + \cdots + (1-p)^k + \cdots \right] \quad (\text{A12})$$

ですが、[]内は等比級数になっていますから、

$$\begin{aligned} E(X) - (1-p)E(X) &= pE(X) = p \cdot \frac{1}{1 - (1-p)} \\ E(X) &= \frac{1}{p} \end{aligned} \quad (\text{A13})$$

となります。同様に、(5)式から

$$\begin{aligned} E(X^2) &= \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} \\ &= p \left[1^2 \cdot (1-p)^0 + 2^2 \cdot (1-p)^1 + 3^2 \cdot (1-p)^2 + \cdots + (k+1)^2(1-p)^k + \cdots \right] \end{aligned} \quad (\text{A14})$$

であり、

$$(1-p)E(X^2) = p \left[1^2 \cdot (1-p)^1 + 2^2 \cdot (1-p)^2 + 3^2 \cdot (1-p)^3 + \cdots + k^2(1-p)^k + \cdots \right] \quad (\text{A15})$$

³解析学の参考書を見てください。

ですから,

$$\begin{aligned} E(X^2) - (1-p)E(X^2) &= p \left[(1^2 - 0^2)(1-p)^0 + (2^2 - 1^2)(1-p)^1 + \cdots + ((k+1)^2 - k^2)(1-p)^k + \cdots \right] \\ E(X^2) &= \sum_{x=0}^{\infty} (2k+1)(1-p)^k \end{aligned} \quad (\text{A16})$$

となります. よって

$$\begin{aligned} E(X^2) &= (2 \times 0 + 1)(1-p)^0 + (2 \times 1 + 1)(1-p)^1 + \cdots + (2 \times k + 1)(1-p)^k + \cdots \\ (1-p)E(X^2) &= (2 \times 0 + 1)(1-p)^1 + (2 \times 1 + 1)(1-p)^2 + \cdots + (2 \times k + 1)(1-p)^{k+1} + \cdots \end{aligned} \quad (\text{A17})$$

ですから,

$$\begin{aligned} E(X^2) - (1-p)E(X^2) &= (1-p)^0 + 2(1-p)^1 + 2(1-p)^2 + \cdots + 2(1-p)^k + \cdots \\ pE(X^2) &= 1 + 2 \cdot \frac{1-p}{1-(1-p)} \\ E(X^2) &= \frac{2-p}{p^2} \end{aligned} \quad (\text{A18})$$

となり,

$$V(X) = E(X^2) - \{E(X)\}^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2} \quad (\text{A19})$$

が得られます.