

MAT120：数量的推論 第12講ハンドアウト 二項実験

前回の講義では、条件付き確率と決定木を学び、コイン投げの結果の数え上げも始めました。今回の講義では、次の大きな道具である二項実験を作っていきます。二つの結果しかもたない実験を何度も繰り返すとき、なぜパスカルの三角形や組合せ $C_{n,r}$ が自然に現れるのかを見ていきます。また、これを二項分布という考え方や、より一般の確率分布という考え方にも結びつけます。この講義のあとには、現実世界のさまざまな「成功回数を数える」状況で、二項の考え方を使えるようになります。

1 導入

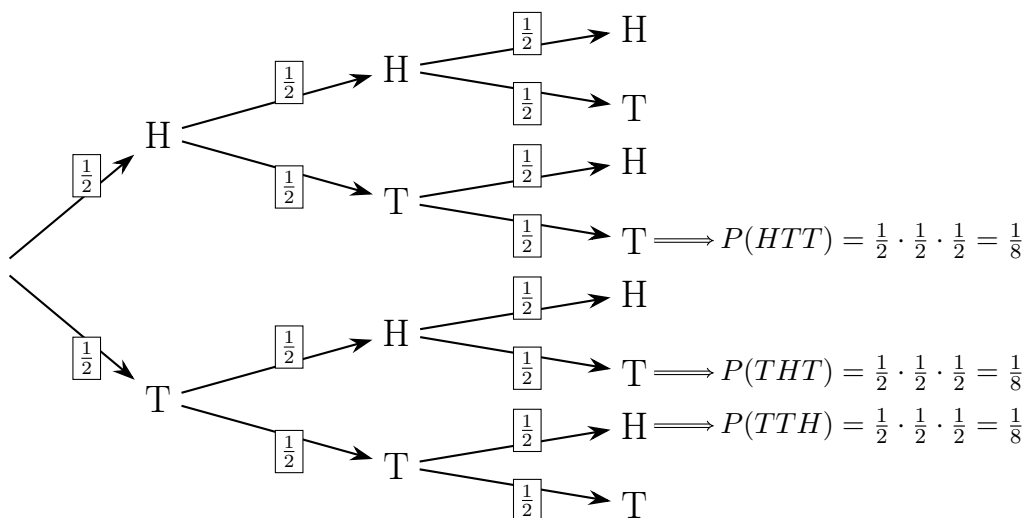
1.1 3回の投げで2回裏が出る確率を計算する

前回の講義では、複数回のコイン投げに関する確率を学びました。その中で、3回のコイン投げで何回裏が出るかを数えようとする、ある種の構造が現れることに気づきました。3のような小さい数であれば、可能な結果をすべて書き出してしまえます。

- コインを1回投げると、結果は2通りあります。H または T です。
- コインを2回投げると、結果は4通りあります。HH, HT, TH, TT です。
- コインを3回投げると、結果は8通りあります（これが標本空間全体です）。

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

もし $P(\text{ちょうど2回}T)$ のような確率を計算したければ、いくつか方法があります。一つの方法は、(第11講のように) 決定木を使って、各有利な結果の確率を計算することです。



そのあと、これらの個別の結果を足し合わせれば、複合事象の確率が得られます。これら3つの結果は互いに排反なので、それぞれの確率を足してよいのです。

$$P(\text{ちょうど2回}T) = P(HTT \text{ または } THT \text{ または } TTH) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}.$$

あるいは、第10講で扱ったような、もっと単純な複合事象の計算でも同じ答えにたどりつけます。この状況ではすべての結果が等確率であることが分かるので、単に次の公式を使ってもよいのです。

$$P(\text{事象}) = \frac{\text{その事象に有利な結果の数}}{\text{結果の総数}}.$$

この場合、ちょうど2回裏を含む結果は HTT, THT, TTH の3つなので、

$$P(\text{ちょうど2回}T) = \frac{\text{ちょうど2回}T \text{ を含む結果の数}}{\text{結果の総数}} = \frac{3}{8}.$$

2 二項実験

2.1 例1：公平なコインを10回投げる

公平なコインを1枚持っていて、それを10回続けて投げるとしましょう。ちょうど5回裏が出る確率はいくらでしょうか。

解答

前の節に基づけば、 $P(\text{ちょうど5回裏})$ は

$$P(\text{ちょうど5回裏}) = \frac{C_{10,5}}{2^{10}}.$$

という公式で計算できます。もちろん通常の公式 $C_{n,r} = \frac{n!}{r!(n-r)!}$ を使ってもよいのですが、流れをそろえるために、ここではパスカルの三角形を使うことにしましょう。パスカルの三角形の10行目は

$$1 \quad 10 \quad 45 \quad 120 \quad 210 \quad 252 \quad 210 \quad 120 \quad 45 \quad 10 \quad 1$$

です。したがって、 $r=5$ に対応する項（パスカルの三角形は0から数え始めることを忘れないでください）を見ると、 $C_{10,5} = 252$ です。結果の総数は $2^{10} = 1024$ です。よって

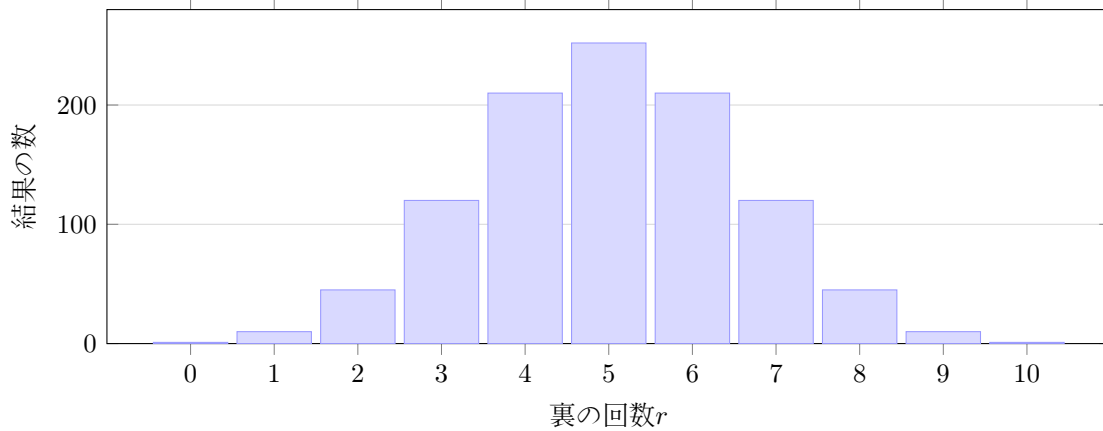
$$P(5回裏) = \frac{C_{10,5}}{2^{10}} = \frac{252}{1024} \approx 0.246 \approx 24.6\%.$$

ここで、10回のコイン投げの標本空間の構造を少し探ってみましょう。上の議論から、この標本空間には1024個の要素があることが分かっています。それらはHHHHHHHHHHからTTTTTTTTTTまで、その間のあらゆるものを含みます。この大きな空間をうまく見渡す自然な方法は、裏が何回含まれているかによって結果をまとめることです。² 標本空間を理解するための便利な方法は、含まれる裏の回数で結果を分類することです。パスカルの三角形の10行目に基づくと、1024個の結果からなる標本空間は次のように整理できます。

裏の回数 r	0	1	2	3	4	5	6	7	8	9	10
結果の数	1	10	45	120	210	252	210	120	45	10	1

この表は、次のような棒グラフとして視覚化できます。

²もちろん、表の回数でまとめてもまったく同じことができます。ただ、ここまでずっと裏を使ってきたので、このまま裏で統一します。

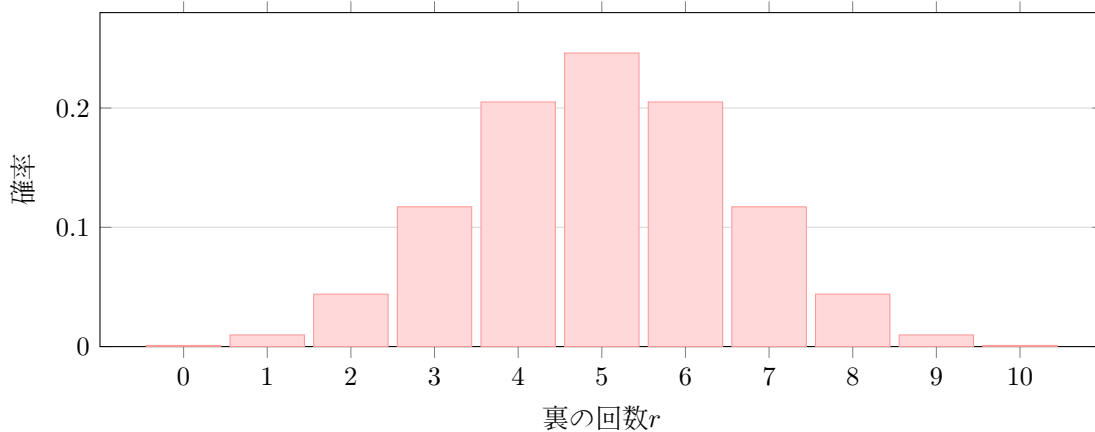


ここで、 x 軸は離散的な個数を表しています。つまり、あり得る裏の回数です。 y 軸は、それに対応する結果の数です。したがって、ある意味でこのグラフは、パスカルの三角形の10 行目を視覚化したものにすぎません。

上の議論から、この系の確率は $C_{10,r}$ を $2^{10} = 1024$ で割ることで計算できると分かっています。そこで実際に割り算を行い、その結果を図にすると、この系の結果全体のあいだに確率そのものがどのように分配されているかを見ることができます。

r 回裏	0	1	2	3	4	5	6	7	8	9	10
確率 (分数)	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$
近似パーセント	0.1%	1.0%	4.4%	11.7%	20.5%	24.6%	20.5%	11.7%	4.4%	1.0%	0.1%

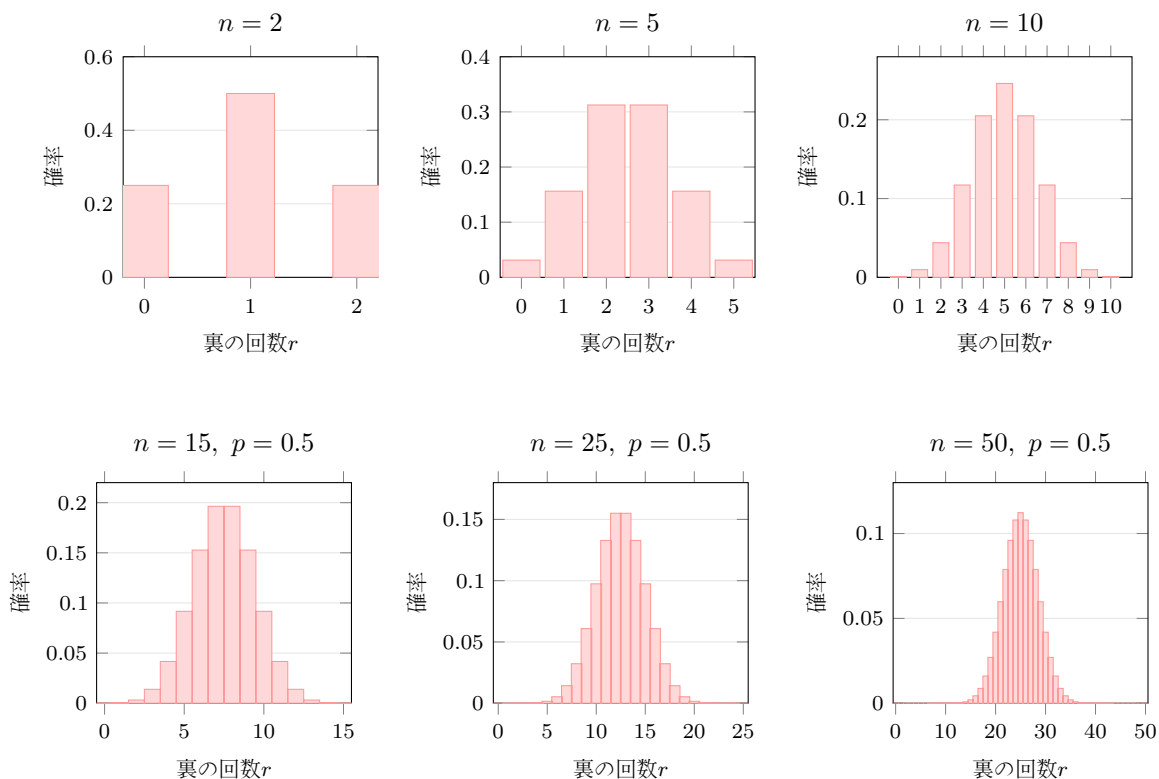
先ほど得た24.6% という値に注目してください。この表を用いると、この確率分布を次のように図示できます。



形が、先ほどの棒グラフとまったく同じ であることに注目してください。変わったのは y 軸の尺度だけです。この特別なグラフには名前があります。これを二項分布 といいます。

2.2 他の n に対する二項分布

一般に、コイン投げの総回数 n の値を変えれば、それに応じた二項分布を作ることができます。これらのグラフの形は、コイン投げの非常に多くの可能な結果のあいだで、確率がどのように分配されているかを表します。その形はパスカルの三角形の n 行目から生じ、それらの数を 2^n で割ると確率になります。一般に、コイン投げをどんどん続けて n が大きくなると、 r の取り得る値も増え、分布はより滑らかになり、真ん中付近 ($r = n/2$ の近く) により集中するようになります。この「滑らかさ」については、第14講で詳しく扱います。



2.3 例2：偏ったコインを投げる

今度は、何らかの細工がされたコインを持っているとしましょう。その結果、表が出る確率と裏が出る確率は、もはや等しくありません。³ この「偏った」コインについて、

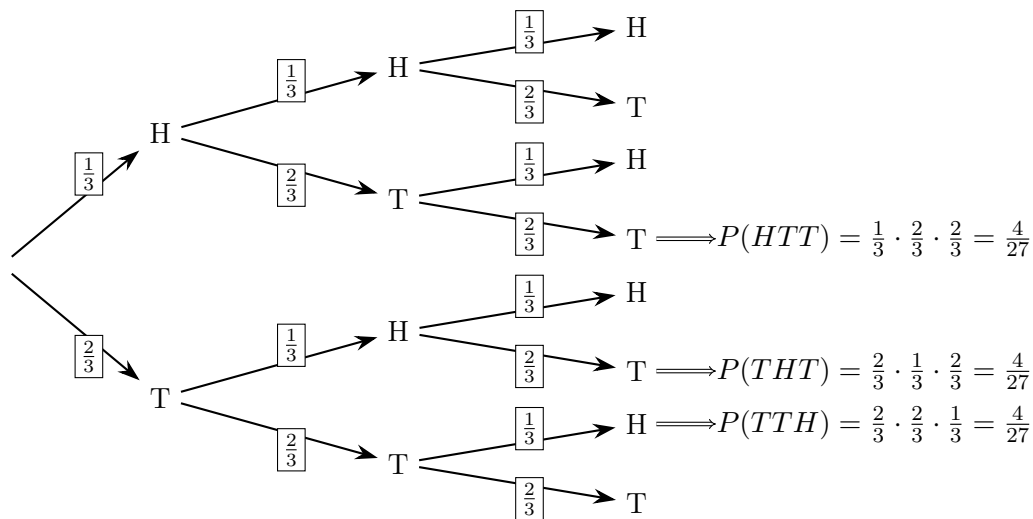
$$P(H) = \frac{1}{3}, \quad P(T) = \frac{2}{3}$$

であると仮定しましょう。問いは、ちょうど2回裏が出る確率はいくらか、です。

上で扱った公平なコインの場合には、8個の結果がすべて等確率だったので、有利な結果の数 (3個ありました) を数えて8で割れば十分でした。しかし、この偏ったコインでは、すべての結果が等確率ではありません。したがって、有利な結果を数えて8で割ることはできません。そ

³現実にこれをどう実現するかは実は私にも完全には分かりませんが、コインを少し曲げて平らでなくし、弓なりへのようにすれば、偏ったコインになるのではないかと思います。片側の表面積が大きくなるので、不公平になりそうです。

の代わりに、決定木の方法を使います。可能な結果の空間そのものは前と同じですが、木の各枝につく確率が変わります。



ちょうど2回裏というのは、 $H T T$, $T H T$, $T T H$ という結果に対応します。これらは互いに排反な事象なので、求める確率はそれらを足し合わせれば得られます。

$$P(\text{ちょうど2回裏}) = P(H T T \text{ または } T H T \text{ または } T T H) = \frac{4}{27} + \frac{4}{27} + \frac{4}{27} = \frac{12}{27}.$$

ここで、各結果の確率計算は実質的に同じであることに注意してください。 $P(H T T)$ でも $P(T H T)$ でも $P(T T H)$ でも、 $P(T) = \frac{2}{3}$ を2回掛け、 $P(H) = \frac{1}{3}$ を1回掛けているだけです。実は、この計算の形はさらに一般の確率にも使えます。 n 回の投げに対しては

$$P(r \text{ 個の } T) = C_{n,r} \left(\frac{2}{3}\right)^r \left(\frac{1}{3}\right)^{n-r}$$

となります。

さらに一般には、裏や表の確率を別の値に変えたとしても、 n 回の投げで r 回裏が出る確率は

$$P(n \text{ 回の投げで } r \text{ 個の } T) = C_{n,r} (P(T))^r (P(H))^{n-r}$$

という公式で計算できます。ここで $(P(T))^r (P(H))^{n-r}$ は、 r 回のTと $(n-r)$ 回のHをもつ特定の並び方の確率であり、 $C_{n,r}$ はそのような並び方が全部で何通りあるかを数えているのです。

3 一般理論

3.1 確率変数

ある変数 X が確率変数と呼ばれるのは、実験や観測において X が取る値がランダムな結果によって決まるときです。主に2種類あります。

- 離散型確率変数は、有限個の値、あるいは可算個の値しか取りません。
- 連続型確率変数は、ある区間の中の無数の値を取り得ます。

演習：離散か連続か

次の各確率変数を、離散型か連続型かに分類しなさい。

- 選ばれた学生が秋学期の登録を終えるまでにかかる時間。
- ランダムに選ばれたある日に、選ばれた学生が受け取るテキストメッセージの本数。
- 電気自動車が満充電で走れる距離（マイル）。
- ある地区の登録有権者から50人の無作為標本を取り、そのうち前回の郡選挙で投票した人数。

解答

- 連続型。
- 離散型。
- 連続型。
- 離散型。

3.2 確率分布

確率分布とは、離散型確率変数のそれぞれの異なる値に確率を割り当てるもの、あるいは連続型確率変数の各区間に確率を割り当てるもののことです。

コインを10回投げる実験では、確率変数 X は裏の回数 r を数えます。 X はランダムに

$$\{0, 1, 2, 3, \dots, 9, 10\}$$

のいずれかの値を取ります。一般に、 n 回のコイン投げにおける裏の回数を数える離散型確率変数 X は、二項分布の x 軸に描かれる値そのものです。

確率分布とは、簡単に言えば、

- 確率変数 X の取り得る値を x 軸に描き、
- それぞれの値に対応する確率を y 軸に描いた

棒グラフのことです。

3.3 二項実験の定義

第2節では、コイン投げの2つのバージョンを見ました。一つは公平なコイン、もう一つは、2つの可能な結果の確率が異なるという意味での「不公平な」コインでした。実は、これらはどちらも二項実験と呼ばれる、より一般的な過程の例です。二項実験には重要な特徴が4つあります。

定義：二項実験

二項実験は次の性質をもちます。

- (B1) 試行回数が固定されている。この回数を n とする。
- (B2) n 回の試行は独立であり、同一条件のもとで繰り返される。
- (B3) 各試行には2つの結果しかない。成功(S)と失敗(F)である。
- (B4) 各試行で、成功の確率は常に同じである。その確率を p 、失敗の確率を q とする。各試行は成功か失敗のどちらかなので、 $p + q = 1$ 、したがって $q = 1 - p$ である。

言い換えれば、二項実験とは、各試行が2つの結果しかもたない多段階の実験です。二項実験における中心的な問題は、 n 回の試行のうち r 回成功する確率を求めることです。仮定(B2)は、こうした連続した確率の計算を簡単にするために使われます。

3.3.1 例：公平なコイン投げ

前節の最初の例では、

- (B1) n はコインを投げる回数でした。
- (B2) コイン投げどうしは独立です。実際、前の結果とはまったく関係していません（物理的に考えてもそうです）。
- (B3) 成功は「裏が出た」と定義し、失敗は「裏が出なかった」と定義しました。
- (B4) 公平なコインでは、 $p = P(T) = 0.5$ であり、 $q = 0.5$ です。

この実験に関する確率を求めるという中心問題に対して、まさにこれまでの議論がそのまま当てはまっていました。公平なコインについてはすでに

$$P(n \text{ 回の投げで } r \text{ 回裏}) = \frac{C_{n,r}}{2^n}$$

であることを見ました。

不公平なコインについても、二項実験としての設定は本質的には同じであり、ただ確率を $p = \frac{2}{3}$ 、 $q = \frac{1}{3}$ に変えただけです。そのとき、 r のさまざまな値に対する確率は

$$P(r \text{ 個の } T) = C_{n,r} \left(\frac{2}{3}\right)^r \left(\frac{1}{3}\right)^{n-r} = C_{n,r} \cdot P(T)^r \cdot P(H)^{n-r}$$

という公式で計算できることを見ました。

3.4 二項確率の一般公式

二項確率の一般公式

二項実験のパラメータが $n, p, q = 1 - p$ であるとき、ちょうど r 回成功する確率は

$$P(n \text{ 回の試行で } r \text{ 回成功}) = C_{n,r} p^r q^{n-r}$$

で与えられます。ここで、

- (1) p^r は、ある特定の順序で成功が r 回起こる確率です。
- (2) q^{n-r} は、同じその順序で失敗が $(n - r)$ 回起こる確率です。
- (3) $p^r q^{n-r}$ は、 r 回の成功と $(n - r)$ 回の失敗をもつ一つの特定の並び方の確率です。
- (4) $C_{n,r}$ は、 r 回の成功と $(n - r)$ 回の失敗をもつ異なる並び方が何通りあるかを数えています。

3.5 p は分布に何をするのか

一般に、

$$P(n \text{ 回の試行で } r \text{ 回成功}) = C_{n,r} p^r q^{n-r}$$

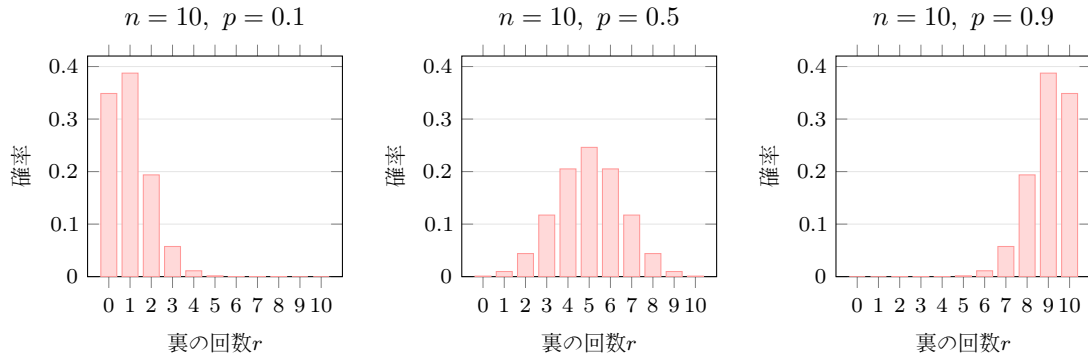
という公式を使えば、任意の二項実験に対応する二項分布を記述できます。単純に言えば、 x 軸には確率変数の取り得る値 (0 から n まで) を置き、 y 軸には上の公式の出力を置きます。

公平なコインの場合には、成功確率 $p = 0.5$ であり、これは q とちょうど等しいです。このため、得られる二項分布は完全に対称になります。第2節のグラフでも、そのことが確認できます。成功確率を $p = 0.5$ からずらしていくと、この対称性は失われます。

イメージのために、 $n = 10$ 回の試行からなる二項実験を考えましょう。 $p = 0.5$ の場合には、対応する分布は第2.1節で見た図でした。ここでは、 p を大きくしたり小さくしたりすると何か起こるかを簡単に見てみます。

- p が0に近い、たとえば $p = 0.1$ の場合、各試行で成功が起こる可能性はかなり低くなります。したがって、10回の試行のうち7回、8回、9回、あるいは10回も成功する確率はきわめて低くなります。たいていは、10回やっても1回か2回しか成功しないでしょう。その結果、このような系に対応する二項分布では、確率の大部分が x 軸の左側、つまり0,1,2あたりに集まります。
- p が1に近い、たとえば $p = 0.9$ の場合には、今度は成功が起こることがきわめて起こりやすくなります。したがって、10回の試行で成功が1回か2回しか起こらないのは非常に不運であり、むしろ7回、8回、9回、あるいは10回程度の成功を期待することになります。二項分布のレベルで見ると、このように成功確率が高いと、分布は右側に押され、確率の大部分は8,9,10のあたりに集中します。

下のグラフは、 $n = 10$ 回の試行に対して成功確率がそれぞれ0.1, 0.5, 0.9 のときの実際の二項分布を表しています。



3.6 例：ルーレット

クラスに22人の学生がいるとしましょう。あなたは一生懸命勉強し、数年後にLUJを無事卒業します。するとDr. O'Connellはあなたをととても誇らしく思い、みんなをカジノに連れて行ってルーレットで遊ばせてくれることにしました。



ルーレットには数字の入ったマスが36個あり、さらに緑のマスが1つあるので、全部で37通りの結果があります。Dr. O'Connellは次のような条件を出します。

ボールが緑のマスに入ったら\$50,000をあげる。それ以外の結果なら\$0だ。

ここで答えたい大きな問いは次です。

演習

ちょうど3人の学生が勝つ確率は何か。

この問いを解くために、第3.4節で述べた二項公式を使います。ただし、計算を始める前に、まずこの状況が本当に二項実験なのかどうかを確認しなければなりません。

- B1：学生は22人いて、それぞれ1回ずつルーレットを回すので、試行回数は $n = 22$ です。

- B2: ルーレットが公平であると仮定すれば, 各試行はたしかに互いに独立です。
- B3: 各試行には2つの結果しかありません。

成功 = 緑に入る, 失敗 = それ以外に入る.

- B4: ルーレットの各マスは等確率で出るとします。マスは全部で37個あり, そのうち緑は1個だけなので, 成功確率と失敗確率は

$$p = P(\text{成功}) = \frac{1}{37}, \quad q = P(\text{失敗}) = 1 - p = \frac{36}{37}$$

です。

これでこの状況を二項実験として扱えることが確認できたので, 本題を解くことができます。

解答

二項公式

$$P(n \text{ 回の試行で } r \text{ 回成功}) = C_{n,r} p^r q^{n-r}$$

を使います。ここでは, 学生が22人いるので $n = 22$, ちょうど3人勝つので $r = 3$, さらに $p = \frac{1}{37}$, $q = \frac{36}{37}$ です。これらを公式に代入すると,

$$P(\text{ちょうど3人勝つ}) = C_{22,3} \left(\frac{1}{37}\right)^3 \left(\frac{36}{37}\right)^{19} = \frac{22!}{3! \cdot 19!} \left(\frac{1}{37}\right)^3 \left(\frac{36}{37}\right)^{19}$$

となります。数値的には

$$P(\text{ちょうど3人勝つ}) \approx 0.018 \approx 1.8\%$$

です。

この確率の別の説明もしておきましょう。ここでは22回の独立な試行を行い, そのうち3回成功, 19回失敗を求めています。そのような結果の一つの具体例として,

$$S, S, S, F, F, \dots, F,$$

すなわち最初の3人が成功し, その後の19人がすべて失敗する, という並びを考えることができます。仮定(B2)により, これらの試行はすべて互いに独立です。したがって, この一つの結果の確率は, 各結果の確率をただ掛け合わせることで計算できます。すなわち

$$P(S \text{ のあと } S \text{ のあと } S \text{ のあと } F \text{ のあと } \dots \text{ のあと } F) = \frac{1}{37} \cdot \frac{1}{37} \cdot \frac{1}{37} \cdot \frac{36}{37} \cdots \frac{36}{37},$$

であり, これは $p^3 q^{19}$ と書くのと同じです。これは, 3回成功し19回失敗する一つの特定の並び方の確率です。しかし, 22回の試行の中に3回の成功を置く方法は $C_{22,3}$ 通りあります。これらの列は互いに排反な結果なので, その確率を全部足し合わせると $C_{22,3} p^3 q^{19}$ が得られます。

$$P(\text{ちょうど3人勝つ}) = \underbrace{p^3 q^{19} + p^3 q^{19} + \dots + p^3 q^{19}}_{C_{22,3} \text{ 回}} = C_{22,3} \cdot p^3 q^{19},$$

となり, これは先ほどの公式と同じです。

4 演習

4.1 演習1：サイコロを振る

標準的な6面サイコロを3回振るとします。

「成功」を「6が出る」と定義しましょう。すると、各試行には2つの結果しかなく（6が出るか、6が出ないか）、各試行は独立であり、成功確率は一定です。

$$n = 3, \quad p = P(6 \text{ が出る}) = \frac{1}{6}, \quad q = \frac{5}{6}.$$

問題

ちょうど2回6が出る確率は何か。

解答

ちょうど2回6が出るためには、 $n = 3$ 回の試行のうち $r = 2$ 回成功すればよいです。二項公式より

$$P(r \text{ 回成功}) = C_{n,r} p^r q^{n-r}$$

です。したがって

$$P(\text{ちょうど2回6}) = C_{3,2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) = 3 \cdot \frac{1}{36} \cdot \frac{5}{6} = \frac{15}{216} = \frac{5}{72} \approx 6.9\%.$$

言葉で言えば、3回のうちどの2回か6になるかを選ぶ方法は $C_{3,2} = 3$ 通りあり、そのそれぞれの並び方の確率は $\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$ である、ということです。

4.2 演習2：バスケットボールをする猫

ある猫がフリースローを打っています。猫はバスケットボールがあまり上手ではないので、シュートが入る確率は $p = 0.5$ だとしましょう。猫は6本シュートを打ちます。

「成功」を「シュートが入ること」とします。すると

$$n = 6, \quad p = 0.5, \quad q = 0.5$$

です。

問題

- (a) 猫がちょうど4本決める確率は何か。
- (b) 猫が少なくとも4本決める確率は何か。

解答

X を入った本数とします。

(a) ちょうど4本入るといのは、 $n = 6$ 回の試行のうち $r = 4$ 回成功することです。

$$P(X = 4) = C_{6,4}(0.5)^4(0.5)^2 = C_{6,4}(0.5)^6 = \frac{C_{6,4}}{64} = \frac{15}{64} \approx 23.4\%.$$

(b) 「少なくとも4本」というのは、4本または5本または6本入ることを意味するので、これらの排反な場合を足し合わせます。

$$\begin{aligned} P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= \frac{C_{6,4}}{64} + \frac{C_{6,5}}{64} + \frac{C_{6,6}}{64} = \frac{15 + 6 + 1}{64} = \frac{22}{64} = \frac{11}{32} \approx 34.4\%. \end{aligned}$$