確率論と統計学

さあ進もう。神々の示現と確率・統計が呼んでいる世界へ。

賽は投げられた。

緑川章一

目 次

第1章　確率論の歴史

サイコロゲームの歴史は古く、人々は何千年も前からサイコロを転がして遊んでいたようである。サイコロの起源は、距骨と呼ばれる鹿や羊の踵の骨で、これを転がすと、区別のつく 4 つの面が出やすい構造になっている。サイコロは漢字で『骰子』とも書くが、この『骰』の字は骨を投げることを意味している。

古代ローマ人もサイコロ遊びが大好きだったようである。紀元前 100 年頃の政治家でもあり軍人でもあったユリウス・カエサルは、政敵ポンペイウスを倒す決心をしてルビコン川を渡るとき、「賽は投げられた。」と叫んだと伝えられている。この言葉の意味するところは、「自分自身の進むべき道は決まった。それが吉と出るか凶と出るかは、神のみぞ知る。」といったところであろうか。ローマに進軍したカエサルは、内乱に勝利し、独裁者としての地位を築いてゆく。

サイコロの魅力は、面の出方がでたらめで事前に予測できない点にある。でたらめは、当て字で『出鱈目』と書くが、この目はサイコロの目ではないかと言われている。世の中には、事前予測ができると言う人もいないわけではないが、そのような人は「イカサマ師」と呼ばれている。そもそも「イカサマ」とは、いかにもその通りと確信をもって言うときに使われる言葉である。この複雑な世の中にあって、確信をもって未来を予測できるのは、ペテン師以外にありえない。

14 世紀から 16 世紀のヨーロッパは、中世の暗黒時代から脱して、ルネサンスを迎える。その時代を生きたイタリアの医師で占星術師で数学者で賭博者でもあったジェロラモ・カルダーノ (1501-1576) は、まさにルネッサンスの寵児であった。彼は、1545 年に数学書『アルス・マグナ』(偉大なる術) を著し、そこに 3 次方程式の解法を載せたのであった。そのために、3 次方程式の解法は、カルダーノの公式と呼ばれることになった。実は、3 次小定式の解法はカルダノによってなされたものではない。彼によれば、まず最初に、デル・フェロによって発見され、その後、ニコロ・フォンタナ（タルタニア）によって独立に発見されたそうである。彼はまた、『偶然のゲーム』という本を著した。ここには、偶然を数学的に取り扱おうとしていて、確率の概念の萌芽が見られるとのことである。

確率論の研究の始まりは、2 人のフランス人、ブレーズ・パスカル (1623-1662) とピエール・ド・フェルマー (1601-1665) のサイコロ遊びについての文通からである。パスカルは、『パンセ』(随想録) を著した人で、「人間は考える葦である」、「ク

レオパトラの鼻がもう少し低かったら世界の歴史は変わっていただろう」などの名文句を残している。また、物理学では、流体の圧力について「パスカルの原理」を発見した。最近では、天気予報でおなじみの名前である。フェルマーは、大学で法律を勉強し弁護士となり、トゥールーズの役人となった。ディオファントス著作の『整数論』の余白に、いわゆる「フェルマーの最終定理」を書き込んだ。これは、$n \geqq 3$ の整数の場合、$x^n + y^n = z^n$ を満たす 0 を含まない整数 $x,\ y,\ z$ の組は存在しないというものである。（$n = 2$ の場合は、ピタゴラスの定理である。）それに続けて、彼は、「驚くべき証明を発見したが、余白が狭すぎる」と書き残した。これが、後代の数学者を悩まし続けることになったが、1994 年にイギリスの数学者アンドリュー・ワイルズによって完全な証明が与えられた。

フランスの貴族シュバリエ・ド・メレは、賭け事が大好きだった。ある時、サイコロ賭博で大損をした。そこで、なぜ負けたのかを知り合いのパスカルに質問した。パスカルは、当代きっての数学者であるフェルマーに助けを求めた。こうして、2 人の間で文通が始まり、順列と組合せの数学が発展し、パスカルは、パスカルの三角形などに名を留めることになった。

オランダのクリスチャン・ホイヘンス (1629-1695) は、物理学者として有名であるが、数学者としても傑出していたようである。パリを訪れたときに、彼は、パスカルとフェルマーの研究について耳にした。確率論に興味を持った彼は、世界初の確率論の本を書いて 1657 年に出版した。

17 世紀から 18 世紀にかけて、スイスのベルヌーイ家は、数学と物理学において優れた業績を上げた人物を輩出している。ドイツにおける音楽家バッハの家系とともに、天才の家系として有名である。この一家の中で最初に活躍してヤコブ・ベルヌーイ (1654-1705) は、弟のヨハンとともにライプニッツとも親交があり、微分積分学の発展に大きく貢献した。ホイヘンスの著書に感銘を受けたヤコブは、20 年近く確率論と格闘し、『大数の法則』を導き、確率論における彼の名を不朽なものにした。また、結果が「成功か失敗」のように 2 つのうちの 1 つである確率分布を『ベルヌーイ分布』と呼ぶが、これは、ヤコブ・ベルヌーイにちなんだものである。

フランス人のド・モアブル (1667-1754) は、宗教的な迫害を逃れて、20 代の後半からずっとイギリスで暮らした。大学などの教育職に就くことを望んでいたが、叶えられずに家庭教師なのでアルバイトで糊口をしのいだ。彼は、ド・モアブルの定理または公式と呼ばれる関係式

$$e^{i\theta} = \cos\theta + i\sin\theta$$

を導いたことで有名である。確率論においては、『正規分布 (normal distribution)』の発見者として知られる。正規分布は、教会の釣鐘のような形をしているので、ベル・カーブとも呼ばれるが、ガウス分布、誤差分布などと呼ばれるも多い。これは、カール・フリードリッヒ・ガウス (1777-1855) が誤差論のなかで詳しく研究したためだと言われている。

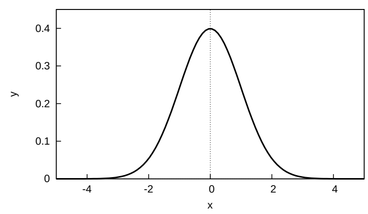
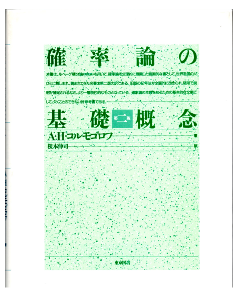


図 1.1: 標準正規分布

イギリスの非国教会派の牧師で数学者であったトマス・ベイズ (1702-1761) は、今日『ベイズの定理』と呼ばれる原因の確率を求める公式を導いた。19 世紀には、ルジャンドル (1752-1833) は、彗星の軌道を決定するために、最小 2 乗法を考え出した。

確率の概念が、ようやくまとまったのは 19 世紀の初頭のことであった。フランスの数学者で天文学者でもあったピエール・ド・シモン・ラプラス (1749-1827) が、大著『確率の解析的基礎』(1814) を著し、ここに古典確率論は完成した。ラプラスの生きた時代は、政治的な動乱期にあった。フランス革命とそれに続くナポレオンの帝政期を経て、王政復古と政治体制は目まぐるしく変化した。多くの科学者が、困難に巻き込まれ研究を続けるのが難しい時代であった。代数方程式論を完結させたエヴァリスト・ガロアもこの時代の人である。ガロアは、決闘によってわずかに 20 歳と 7 か月でこの世を去った。ラプラスは、遊泳術に長けた人で、後世、無節操と批判を浴びることになる。

20 世紀初頭に、フランスの数学者エミール・ボレル (1871-1956) やアンリ・レオン・ルベーグ (1875-1941) によって測度論が考案された。これは、測度（長さ、面積、体積など) を数学的に厳密に扱う数学である。アンドレイ・ニコラエヴィッチ・コルモゴロフ (1903-1987) は、ロシアが帝国だった時に生まれ、ソビエト社会主義共和国連邦と呼ばれた時代に人生の大半を過ごした。ロシア連邦となったのは、彼の死後のことである。彼は、測度論に基づいて確率論を再構成し、1933 年に『確率論の基礎概念』を表す。これにより、確率論の公理化が完成する。



第2章　確率の基礎

2.1　確率の公理

起こりうる全ての事象の集合を標本空間、または、全事象と言い、Ω で表す。Ω の部分集合を要素とする集合を $\mathcal{F}$ とする。$\mathcal{F}$ の要素 A に対して実数関数 P(A) を定義し、この値 P(A) を事象 A の確率と言う。それは次の性質を満たす。

1. $0 \leqq \mathrm{P(A)}$
2. $\mathrm{P}(\Omega) = 1$
3. 互いに共通部分を持たない部分集合 $\mathrm{A_1},\ \mathrm{A_2},\ \cdots,\ \mathrm{A_n}$ に対して、

   $$\mathrm{P(A_1 \cup A_2 \cup \ \cdots \ \cup\ A_n) = P(A_1) + P(A_2) + \ \cdots\ + P(A_n)}$$

   が成り立つ。

標本空間 Ω の中の、これ以上に分けられない事象を**根元事象**と言う。

$\mathrm{P(A)} \leqq 1$ は、上の性質から以下のようにして導かれる。$\mathcal{F}$ の要素 A の補集合を $\mathrm{A^c}$ とすると、

$$\Omega = \mathrm{A \cup A^c} \text{ かつ } \quad \mathrm{A \cap A^c} = \emptyset$$

である。ゆえに、$\mathrm{P(A) + P(A^c)} = \mathrm{P}(\Omega) = 1$. すなわち、
$\mathrm{P(A) = 1 - P(A^c)}$. ところが、$\mathrm{P(A^c)} \geqq 0$ より、$\mathrm{P(A)} \leqq 1$ である。

【例】

サイコロの目の集合の標本空間は、

$$\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$

である。根元事象とは、サイコロの各々の目の集合

$$\{1\},\ \{2\},\ \{3\},\ \{4\},\ \{5\},\ \{6\}$$

が根元事象である。奇数 (odd) の目の集合

$$A_o = \{1,\ 3,\ 5\}$$

や、偶数 (even) の目の集合

$$A_e = \{2,\ 4,\ 6\}$$

などを複合事象と言う。標本空間 Ω の部分集合の総数は、空集合 $\emptyset$ と Ω 自身も含めると、$2^6 = 64$ である。

余事象の確率

確率の総和は 1 なので、ある事象 A の起こる確率 $\mathrm{P}(A)$ とその余事象 A^c の起こる確率 $\mathrm{P}(A^c)$ の和は 1 である。

$$\mathrm{P}(A) + \mathrm{P}(A^c) = 1$$

確率 $\mathrm{P}(A)$ を求めたいが計算が複雑な場合には、その余事象の起こる確率 $\mathrm{P}(A^c)$ を求めて、その値を 1 から引いてやれば良い。

確率論の起源は、フランスの貴族シュヴァリエ・ド・メレが、サイコロ賭博で大損をして、パスカルに相談したことから始まったと言われている。そのド・メレの行った賭けとは、
「2 つのサイコロを 24 回振った場合、少なくとも 1 回は 6 のゾロ目が出るか？」
というもので、彼は出る方に賭けたのだった。この確率を直接計算しようとすると、場合分けが面倒である。24 回中 1 回も 6 のゾロ目が出ない確率を計算する方が簡単である。

2 コのサイコロを振った場合、目の出方は全部で $6 \times 6 = 36$ 通りである。そのうち、6 のゾロ目以外のの目の出方は 35 通りである。そこで、その確率は、35/36 となる。これが 24 回続く確率は、

$$\left(\frac{35}{36}\right)^{24} = 0.508596$$

となり、0.5 よりもわずかに大きくなる。したがって、サイコロを 24 回振って、少なくとも 1 回は 6 のゾロ目が出る確率は、

$$1 - \left(\frac{35}{36}\right)^{24} = 0.491404$$

である。

注

$$\left(\frac{35}{36}\right)^{24} = \left(1 - \frac{1}{36}\right)^{24} > 1 - \frac{24}{36} + \frac{24 \cdot 23}{2 \cdot 1} \cdot \frac{1}{36^2} - \frac{24 \cdot 23 \cdot 22}{3 \cdot 2 \cdot 1} \cdot \frac{1}{36^3}$$

$$\therefore \left(\frac{35}{36}\right)^{24} > 1 - 0.6667 + 0.2129 - 0.0434 = 0.5028$$

2.2 条件付確率

一般に、標本空間 Ω の部分集合 A, B が与えられていて、かつ $P(A) \neq 0$ の場合に、

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{2.1}$$

を事象 A が起ったときの事象 B の**条件付確率**と言う。言い換えると、条件付確率 $P(B|A)$ とは A が起ったと分った時に、それが B である確率である。

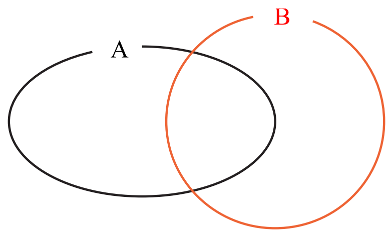


図 2.1: 共通部分と条件付確率：事象 A のもとで事象 B の起こる確率は、$P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ で与えられる。

【例題】

男の子、または女の子の生まれる確率は、ほぼ 1/2 である。実際には男の子の生まれる確率の方が少し大きいが、これは、非常に良い近似である。2 人の子どもがいる場合、2 人とも男の子の場合、上が男の子で下が女の子の場合、2 人とも女の子の場合の 4 通りがある。これらは、全て同等に確からしいので、その確率はそれぞれ 1/4 である。これを、

$$P(\text{男男}) = P(\text{男女}) = P(\text{女男}) = P(\text{女女}) = \frac{1}{4}$$

と、1 番目を最初に 2 番目を後に書こう。

ある家庭には 2 人の子どもがいるが、少なくとも 1 人は女の子だと分かったとしよう。すると、(男男) の組合せは無くなるので、考えられる場合の数は 3 通りとなうるので、 2 人とも女の子の確率は、

$$\mathrm{P}((\text{女女})|\text{ すくなくとも 1 人は女}) = \frac{\mathrm{P}(\text{女女})}{\mathrm{P}(\text{男女}) + \mathrm{P}(\text{女男}) + \mathrm{P}(\text{女女})}$$
$$= \frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{4}+\frac{1}{4}} = \frac{1}{3}$$

となる。今の場合は、考えられる場合の数が 3 通りあって、その内の 1 つが起こる確率なので、答えは 1/3 とやっても問題ないが、それは、3 通りの場合が同様に確からしいためである。

応用

(2.1) 式を、

$$\mathrm{P}(A \cap B) = \mathrm{P}(B|A)\mathrm{P}(A) \tag{2.2}$$

と書こう。集合 A の補集合を A^c と書くと、

$$\mathrm{P}(A^c \cap B) = \mathrm{P}(B|A^c)\mathrm{P}(A^c) \tag{2.3}$$

が成り立つ。

ところが、

$$(A \cap B) \cup (A^c \cap B) = (A \cup A^c) \cap B = B$$

かつ、

$$(A \cap B) \cap (A^c \cap B) = \emptyset$$

なので、

$$\mathrm{P}(B) = \mathrm{P}(A \cap B) + \mathrm{P}(A^c \cap B)$$

が成り立つ。ゆえに、

$$\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A^c)\mathrm{P}(A^c) \tag{2.4}$$

である。

【例題】

100 本のくじの中に 1 本の当りくじがある。1 番目に引く人と 2 番目に引く人が当りくじを引く確率は、それぞれいくらか。100 本の中の 1 本が当りくじだから、どれを引いても当る確率は、1/100 で同じだ、と言ってしまえばそれまでだが、1 番目の人が引くと、くじは 1 本減って、99 本になってしまう。それでも、2 番目に引く人が当る確率は 1/100 だろうか。

1 番目の人が当りくじを引く確率は、$\mathrm{P}(1_{\text{当}}) = 1/100$、外れくじを引く確率は、$\mathrm{P}(1_{\text{外}}) = 99/100$ である。1 番目の人が当りくじを引いた場合に 2 番目の人が当りくじを引く確率を $\mathrm{P}(2_{\text{当}}|1_{\text{当}})$ と書くことにすると、この値は、当然のことながら 0 である。一方、1 番目の人が外れくじを引いた場合に 2 番目の人が当りくじを引く確率は、$\mathrm{P}(2_{\text{当}}|1_{\text{外}}) = 1/99$ である。そこで、2 番目の人が当りくじを引く確率は、

$$\begin{aligned}\mathrm{P}(2_{\text{当}}) &= \mathrm{P}(2_{\text{当}}|1_{\text{当}})\mathrm{P}(1_{\text{当}}) + \mathrm{P}(2_{\text{当}}|1_{\text{外}})\mathrm{P}(1_{\text{外}}) \\ &= 0 \times \frac{1}{100} + \frac{1}{99} \times \frac{99}{100} \\ &= \frac{1}{100}\end{aligned}$$

となり、やはり $\mathrm{P}(2_{\text{当}}) = 1/100$ であることがわかる。

2.3　ベイズの定理

$\{A_1,\ A_2,\ \cdots,\ A_n\}$ を Ω の分割とする。すなわち、

$$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega, \quad \text{かつ} \quad i = j \text{ のとき } A_i \cap A_j = \emptyset$$

である。このとき、

$$\mathrm{P}(B|A_i) = \frac{\mathrm{P}(B \cap A_i)}{\mathrm{P}(A_i)}$$

ゆえに、

$$\mathrm{P}(B|A_i)\mathrm{P}(A_i) = \mathrm{P}(B \cap A_i)$$

一方、

$$B = B \cap \Omega = \sum_{k=1}^{n} B \cap A_k$$

だから、

$$\mathrm{P}(B) = \sum_{k=1}^{n} \mathrm{P}(B \cap A_k) = \sum_{k=1}^{n} \mathrm{P}(B|A_k)\mathrm{P}(A_k) \tag{2.5}$$

が成り立つ。

　また、

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(B \cap A_i)}{\mathrm{P}(B)}$$

である。ここで、右辺の分母 $P(B)$ に (2.5) 式を代入すると、

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(B \cap A_i)}{\displaystyle\sum_{k=1}^{n} \mathrm{P}(B|A_k)\mathrm{P}(A_k)}$$

を得る。これを、**ベイズ (Bayes) の定理**と言う。

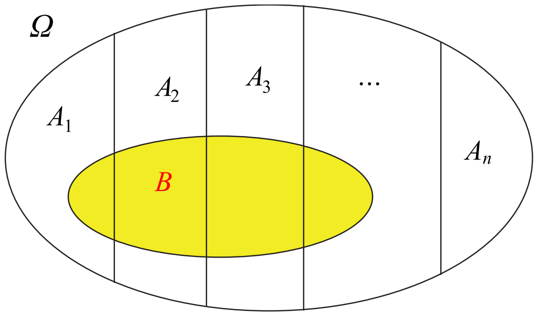


図 2.2: ベイズの定理

ベイズの定理

事象 B は、$\mathrm{P}(B) > 0$ を満たすものとする。$\{A_1, A_2, \cdots, A_n\}$ を標本空間 Ω の分割とする。すなわち、$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$ で、任意の $i \neq j$ について $A_i \cap A_j = \emptyset$ とする。このとき、

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(B|A_i)\mathrm{P}(A_i)}{\sum_{k=1}^{n} \mathrm{P}(B|A_k)\mathrm{P}(A_k)}$$

が成り立つ。$\mathrm{P}(A_i)$ は、事象 B が起る前の確率なので、$\mathrm{P}(A)$ を事前確率と呼ばれる。一方、$\mathrm{P}(A_i|B)$ は、事象 B が起った後の、その原因 A_i についての確率なので、 事後確率と呼ばれる。

特別に、標本空間 Ω が 2 つの集合 $\{A_1,\ A_2\}$ に分割されている場合について考えよう。また、Ω は、2 つの集合 $\{B_1,\ B_2\}$ を含むとする。

$$B_1 = (B_1 \cap A_1) \cup (B_1 \cap A_2). \quad B_2 = (B_2 \cap A_1) \cup (B_2 \cap A_2)$$

ところで、条件付確率の定義から

$$\mathrm{P}(B_1|A_1) = \frac{\mathrm{P}(A_1 \cap\ B_1)}{\mathrm{P}(A_1)}$$

$$\mathrm{P}(B_1|A_2) = \frac{\mathrm{P}(A_2 \cap\ B_l)}{\mathrm{P}(A_2)}$$

が導かれるが、これらを用いると、

$$\mathrm{P}(B_1) = \mathrm{P}(B_1 \cap A_1) + \mathrm{P}(B_1 \cap A_2) = \mathrm{P}(B_1|A_1)\mathrm{P}(A_1) + \mathrm{P}(B_1|A_2)\mathrm{P}(A_2)$$

となる。ところで、$\mathrm{P}(A_1|B_1) = \mathrm{P}(A_1 \cap B_1)/\mathrm{P}(B_1)$ だから、

$$\mathrm{P}(A_1|B_1) = \frac{\mathrm{P}(B_1|A_1)\mathrm{P}(A_1)}{\mathrm{P}(B_1|A_1)\mathrm{P}(A_1) + \mathrm{P}(B_1|A_2)\mathrm{P}(A_2)}$$

とも書ける。同様にして、

$$\begin{aligned}
\mathrm{P}(A_2|B_1) &= \frac{\mathrm{P}(B_1|A_2)\mathrm{P}(A_2)}{\mathrm{P}(B_1|A_1)\mathrm{P}(A_1) + \mathrm{P}(B_1|A_2)\mathrm{P}(A_2)} \\
\mathrm{P}(A_1|B_2) &= \frac{\mathrm{P}(B_2|A_1)\mathrm{P}(A_2)}{\mathrm{P}(B_2|A_1)\mathrm{P}(A_1) + \mathrm{P}(B_2|A_2)\mathrm{P}(A_2)} \\
\mathrm{P}(A_2|B_2) &= \frac{\mathrm{P}(B_2|A_2)\mathrm{P}(A_2)}{\mathrm{P}(B_2|A_1)\mathrm{P}(A_1) + \mathrm{P}(B_2|A_2)\mathrm{P}(A_2)}
\end{aligned}$$

を得る。

これは、図 (2.3) 見ると分かりやすい。右側に事象 A_i $(i = 1,\ 2)$ を、左側に事象 B_j $(j = 1,\ 2)$ を書く。それらを矢印で結んで、その上に、事象 A_i のもとでの事象 B_i の条件付き確率 $\mathrm{P}(B_j|A_i)$ を書き込む。このような図は、情報理論において雑音を含む通信路の解析に良くもちいられ、**通信路線図 (channel diagram)** と呼ばれている。図は、事象 A_1 または A_2 が起った場合に、事象 B_1 または B_2 が生起する場合の確率の流れを表す。事象 A_1 が起る確率を $\mathrm{P}(A_1)$ とすると、A_1 から B_1 への確率の流れは、$\mathrm{P}(B_1|A_1)\mathrm{P}(A_1)$ で与えられる。同様に、事象 A_2 の生起確率を $\mathrm{P}(A_2)$ とすると、A_2 から B_1 への確率の流れは、$\mathrm{P}(B_1|A_2)\mathrm{P}(A_2)$ で与えられる。二つを足すと、確率 $\mathrm{P}(B_1)$ が得られる。

$$\mathrm{P}(B_1) = \mathrm{P}(B_1|A_1)\mathrm{P}(A_1) + \mathrm{P}(B_1|A_2)\mathrm{P}(A_2)$$

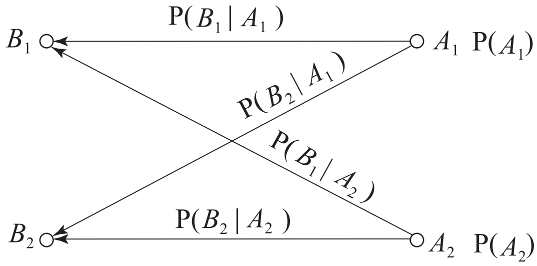


図 2.3: 通信路線図

このうちで、A_1 から流れ込んだ量 $\mathrm{P}(B_1|A_1)\mathrm{P}(A_1)$ の割合が、B_1 が起ったと分った場合にそれが A_1 に起因する確率 $\mathrm{P}(A_1|B_1)$ に他ならない。

B_2 の場合については、

$$\mathrm{P}(B_2) = \mathrm{P}(B_2|A_1)\mathrm{P}(A_1) + \mathrm{P}(B_2|A_2)\mathrm{P}(A_2)$$

が得られ、これも同様な解釈ができる。

【例題 **1**】

100 本のくじの中に 2 本の当りくじがある。A さんが 1 番目に、B さんが 2 番目に 1 本を引くとする。B さんが当りくじを引いたと分かったとき、A さんが当りくじを引いた確率 $\mathrm{P}(A_{当}|B_{当})$ はいくらか。

ベイズの定理を持ち出すまでもなく、B さんが当りくじを引いたので、A さんが当りくじを引くためには、残りの 99 本から当りの 1 本を引かなければならない。ゆえに、$\mathrm{P}(A_{当}|B_{当}) = \dfrac{1}{99}$ である。したがって、この問題にベイズの定理を用いることは、鶏を裂くに牛刀を用いるようなものであるが、敢えてやってみよう。

A さんが、当りくじ、外れくじを引く確率は、各々

$$\mathrm{P}(A_{当}) = \frac{2}{100} = \frac{1}{50}, \qquad \mathrm{P}(A_{外}) = \frac{98}{100} = \frac{49}{50}$$

である。また、A さんが、当りくじ、または、外れくじを引いたと分かっとき、B さんが当りくじを引く確率は、各々、

$$\mathrm{P}(B_{当}|A_{当}) = \frac{1}{99}, \qquad \mathrm{P}(B_{当}|A_{外}) = \frac{2}{99}$$

である。ゆえに、

$$\begin{aligned}\mathrm{P}(A_{当}|B_{当}) &= \frac{\mathrm{P}(B_{当}|A_{当})\mathrm{P}(A_{当})}{\mathrm{P}(B_{当}|A_{当})\mathrm{P}(A_{当})+\mathrm{P}(B_{当}|A_{外})\mathrm{P}(A_{外})} \\ &= \frac{\frac{1}{99}\cdot\frac{1}{50}}{\frac{1}{99}\cdot\frac{1}{50}+\frac{2}{99}\cdot\frac{49}{50}} = \frac{1}{1+2\cdot 49} = \frac{1}{99}\end{aligned}$$

となり、先ほどの結果と一致する。

【例題 **2**】

ある家庭には 2 人の子どもがいる。その家庭を訪問したところ、女の子が現れた。2 人とも女の子である確率はいくらか。

少なくとも 1 人は女の子だから、可能な 3 通りの内の 1 つなので 1/3 だ、と早合点してはいけない。出て来なかったもう 1 人が、女の子である確率は、1/2。ゆえに、 2 人とも女の子である確率は、1/2 である。

この問題をベイズの定理を用いて解いてみよう。

2 人兄弟の性別による分類は 4 通りで、それらはすべて同様に確からしいから、

$$\mathrm{P}(男,男)=\mathrm{P}(男,女)=\mathrm{P}(女,男)=\mathrm{P}(女,女)=\frac{1}{4}$$

である。これら 4 つのケースの各々について、女の子が出てくる条件付き確率を求めると、

$$\begin{aligned}&\mathrm{P}(女出\,|\,男,男)=0, &&\mathrm{P}(女出\,|\,男,女)=\frac{1}{2},\\ &\mathrm{P}(女出\,|\,女,男)=\frac{1}{2}, &&\mathrm{P}(女出\,|\,女,女)=1\end{aligned}$$

女の子が出てくる確率 P(女出) は、

$$\begin{aligned}\mathrm{P}(女出) &= P(女出\,|\,男,男)P(男,男)+P(女出\,|\,男,女)P(男,女)\\ &\quad +P(女出\,|\,女,男)P(女,男)+P(女出\,|\,女,女)P(女,女)\\ &= 0\times\frac{1}{4}+\frac{1}{2}\times\frac{1}{4}+\frac{1}{2}\times\frac{1}{4}+\frac{1}{4}=\frac{1}{2}\end{aligned}$$

ゆえに、女の子が出てきたときに、その家庭が2人とも女の子である確率 P(女, 女 | 女出) は、

$$\mathrm{P}(女,女\,|\,女出)=\frac{\mathrm{P}(女出\,|\,女,女)\mathrm{P}(女,女)}{\mathrm{P}(女出)}=\frac{\frac{1}{4}}{\frac{1}{2}}=\frac{1}{2}$$

となる。

図 2.4: 子どもが 2 人の家庭を訪問して女の子が出てきた場合

【**例題 3**】

(1) いま、0 と 1 を発信する情報源があるとしよう。この情報源が 0 を発信する確率を 0.6、 1 を発信する確率を 0.4 とする。この信号を正確に受信する確率を、0.9、誤って受信する確率を 0.1 とする。受信信号 0 を受け取った場合に、情報源の入力信号が 0 または 1 であった確率はいくらだろうか。

入力信号が 0，1 の確率を

$$\mathrm{P}(0_{in}) = 0.6,\ \mathrm{P}(1_{in}) = 0.4$$

と表わそう。それに対して、受信信号が 0、1 である確率を、それぞれ、$\mathrm{P}(0_{out})$、$\mathrm{P}(1_{out})$ と書こう。信号が正しく伝わる確率は、

$$\mathrm{P}(0_{out}|0_{in}) = \mathrm{P}(1_{out}|1_{in}) = 0.9$$

で、誤って伝わる確率は、

$$\mathrm{P}(0_{out}|1_{in}) = \mathrm{P}(1_{out}|0_{in}) = 0.1$$

と表わされる。0 または 1 を受信する確率は、それぞれ、

$$\begin{aligned}
\mathrm{P}(0_{out}) &= \mathrm{P}(0_{out}|0_{in})\mathrm{P}(0_{in)} + \mathrm{P}(0_{out}|1_{in})\mathrm{P}(1_{in}) \\
&= 0.9\times\ 0.6 + 0.1\times\ 0.4 = 0.58 \\
\mathrm{P}(1_{out}) &= \mathrm{P}(1_{out}|0_{in})\mathrm{P}(0_{in}) + \mathrm{P}(1_{out}|1_{in})\mathrm{P}(1_{in}) \\
&= 0.1\times\ 0.6 + 0.9\times\ 0.4 = 0.42
\end{aligned}$$

と求まる。受信信号 0 を受け取った場合に、入力信号が 0 または 1 であった確率は、それぞれ、

$$\begin{aligned} \mathrm{P}(0_{in}|0_{out}) &= \frac{\mathrm{P}(0_{out}|0_{in})\mathrm{P}(0_{in})}{\mathrm{P}(0_{out})} \\ &= \frac{0.54}{0.58} \simeq 0.931 \\ \mathrm{P}(1_{in}|0_{out}) &= \frac{\mathrm{P}(0_{out}|1_{in})\mathrm{P}(1_{in})}{\mathrm{P}(0_{out})} \\ &= \frac{0.04}{0.58} \simeq 0.069 \end{aligned}$$

となる。

(2) 先ほどと同じように、0 と 1 を発信する情報源があるとする。しかし今度は、入力信号が一方に偏っている場合について考えてみる。例えば、0 を入力する確率が 0.91 で、1 を入力する確率は 0.09 としよう。すなわち、

$$\mathrm{P}(0_{in}) = 0.91, \qquad \mathrm{P}(1_{in}) = 0.09$$

とする。また、信号が正しく伝わる確率と誤って伝わる確率は、先ほどと同じで、

$$\begin{aligned} \mathrm{P}(0_{out}|0_{in}) &= \mathrm{P}(1_{out}|1_{in}) = 0.9 \\ \mathrm{P}(0_{out}|1_{in}) &= P(1_{out}|0_{in}) = 0.1 \end{aligned}$$

とすると、受信信号 0，1 の確率は、それぞれ、

$$\begin{aligned} \mathrm{P}(0_{out}) &= \mathrm{P}(0_{out}|0_{in})\mathrm{P}(0_{in}) + \mathrm{P}(0_{out}|1_{in})\mathrm{P}(1_{in}) \\ &= 0.9 \times 0.91 + 0.1 \times 0.09 \\ &= 0.828 \\ \mathrm{P}(1_{out}) &= \mathrm{P}(1_{out}|0_{in})\mathrm{P}(0_{in}) + \mathrm{P}(1_{out}|1_{in})\mathrm{P}(1_{in}) \\ &= 0.1 \times 0.91 + 0.9 \times 0.09 \\ &= 0.172 \end{aligned}$$

となる。事後確率は、それぞれ、

$$\begin{aligned}
P(0_{in}|0_{out}) &= \frac{P(0_{out}|0_{in})P(0_{in})}{P(0_{out})} = \frac{0.9 \times 0.91}{0.828} \approx 0.989 \\
P(1_{in}|0_{out}) &= \frac{P(0_{out}|1_{in})P(1_{in})}{P(0_{out})} = \frac{0.1 \times 0.09}{0.828} \approx 0.011 \\
P(0_{in}|1_{out}) &= \frac{P(1_{out}|0_{in})P(0_{in})}{P(1_{out})} = \frac{0.1 \times 0.91}{0.172} \approx 0.529 \\
P(1_{in}|1_{out}) &= \frac{P(1_{out}|1_{in})P(1_{in})}{P(1_{out})} = \frac{0.9 \times 0.09}{0.172} \approx 0.471
\end{aligned}$$

となる。すなわち、受信信号が 0 であった場合に入力信号が 0 である確率は 0.989 なので、そのほとんどは 0 を正しく受信したと言うことができる。しかし、1 を受信した場合に入力信号が 1 である確率は 0.471 なので、僅かではあるが、入力信号が 0 であった可能性の方が高くなるのである。これは以外な結果だと思われるかも知れないが、もともと 1 の入力信号は少なかったので、受信信号 1 のうちの半分以上は、0 の入力信号が誤って伝えられたものであることによる。

このような状況は、例えば、非常にまれな病気などを検査で見つける場合などにも起こり得る。罹病率と検査結果の信頼性によっては、たとえ検査結果で陽性反応が出たとしても、それが偽陽性である可能性が高い場合もあり得るのである。

第3章　確率変数と分布

3.1　確率変数

サイコロを振った時の目の出方は6通りあり、これを、

$$\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$

と書いた。サイコロを振って1の目が出る事象を1と表したのである。他の場合 $(2,\ 3,\ \cdots,\ 6)$ についても同様である。このように、事象を数値で表すと何かと便利である。そのためには、事象に数値を対応させる規則を定める必要がある。この規則のことを確率変数と言う。確率変数とは実数の値をとる関数 (集合論で言うところの写像) であることに注意しよう。確率変数は、通常、大文字 X で表す。先ほどのサイコロの例の場合には、

$$x = X(\text{サイコロを振って}\ x\ \text{の目が出る事象})$$

となる。ここで、左辺の小文字 x は1から6までの自然数の値をとる**変数**で、これを**確率変数**と呼ぶ場合もある。

この例のように、確率変数がとびとびの値をとるとき、X は**離散的**であると言う。確率変数 X が値 x をとる確率を $\mathrm{P}(x)$ と表し、これを**確率関数**と言う。確率関数は、また、**確率分布**とも呼ばれる。例えば、サイコロが理想的な場合には、確率分布は、

$$\mathrm{P}(1) = \mathrm{P}(2) = \mathrm{P}(3) = \mathrm{P}(4) = \mathrm{P}(5) = \mathrm{P}(6) = \frac{1}{6}$$

となる。

硬貨を放り投げた場合には、表が出るか裏が出るかのいずれかである。この場合にも、表と裏を数字で表すのが便利であるが、どんな数字を対応させるかは、まったく自由である。そこで、確率変数 X の値を、表が出た場合には1、裏が出た場合には0と決めよう。すなわち、

$$\begin{aligned} 1 &= X(\text{硬貨の表が出る事象}) \\ 0 &= X(\text{硬貨の裏が出る事象}) \end{aligned}$$

である。理想的な場合の確率分布は、

$$\mathrm{P}(1) = \mathrm{P}(0) = \frac{1}{2}$$

となる。

数学者のアンリ・ポアンカレには、毎日買ってきたパンの重さを測り続け、売られているパンの重さが表示されている値よりも軽いことを突き止めた逸話がある。確率変数が、パンの重さのように連続的な値をとる場合には、X は**連続的**であるという。その範囲が微小区間 $x \leqq X \leqq x + dx$ であるときに、

$$\mathrm{P}(x \leqq X \leqq x + dx) = f(x)\,dx$$

と表せるならば、$f(x)$ を**確率密度関数**と言う。確率変数 X が、有限区間 $a \leqq X \leqq b$ である確率 $\mathrm{P}(a \leqq X \leqq b)$ は、

$$\mathrm{P}(a \leqq X \leqq b) = \int_a^b f(x)\,dx$$

で与えられる。

3.2 平均と分散

確率分布が与えられれば、確率規則は完全に決まってしまう。もはや、すべきことは何も無いと思われるかも知れない。しかし、確率のふるまいを大づかみに知りたい場合には、分布を特徴づける簡単な量があると便利である。それが平均と分散である。

平均

確率変数 x の値は、場合毎に異なっている。その代表的な値について知りたい場合に用いられるのが**平均** (mean) である。平均とは、不ぞろいなものを平らに均 (なら) すことであるが、その均し方には色々な方法がある。その中でも、よく用いられるのは、x に P(x) の重みをつけて足したものである。これを、mean の頭文字 m に対応するギリシア文字 μ で表すことにすると、

$$\mu = \sum_x x\mathrm{P}(x) \tag{3.1}$$

となる。これは、また、**期待値** (expectation) と呼ばれることもある。例えば、理想的なサイコロを一回振る場合、1 から 6 までの目の出る確率は、各々1/6 であるので、目の平均は、

$$\mu = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2} = 3.5$$

である。

連続分布の場合の平均は、(3.1) 式において、$\mathrm{P}(x)$ を $f(x)\,dx$ で、和を積分で置き換えることにより求められる。すなわち、

$$\mu = \int x f(x)\,dx$$

で与えられる。

分散

平均とならんで重要なものは、確率変数の平均からの散らばり具合を表す量で、分散 (variance) と呼ばれている。これは σ^2 と表され、確率変数 x の平均 μ からのずれの 2 乗に確率 $\mathrm{P}(x)$ の重みをつけて足したもの、すなわち、

$$\sigma^2 = \sum_x (x-\mu)^2 \mathrm{P}(x) \tag{3.2}$$

で定義される。ここで、$(x-\mu)$ ではなく、$(x-\mu)^2$ に $\mathrm{P}(x)$ を掛けていることに注意しよう。確率変数 x の平均 μ からのずれは、正の場合も負の場合のあるので、$(x-\mu)$ に確率 $\mathrm{P}(x)$ の重みをつけて足したものは、0 となってしまうからである。分散は、$(x-\mu)$ の 2 乗について計算しているので、実際の散らばり具合を知るためには、その平方根をとる必要がある。これを標準偏差 (standard deviation) と呼び、ギリシア文字 σ で表す。すなわち、

$$\sigma = \sqrt{\sigma^2}$$

サイコロの場合の分散は、

$$\sigma^2 = \frac{1}{6}\sum_{x=1}^{6}\left(x-\frac{7}{2}\right)^2 = \frac{35}{12} \simeq 2.9$$

したがって、標準偏差は、

$$\sigma = \sqrt{\frac{35}{12}} \simeq 1.7$$

となる。

分散の式 (3.2) は、$(x-\mu)^2$ を展開することにより、次のように変形することができる。

$$\begin{aligned}
\sigma^2 &= \sum_x (x-\mu)^2 \mathrm{P}(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2)\mathrm{P}(x) \\
&= \sum_x x^2 \mathrm{P}(x) - 2\mu \sum_x x\mathrm{P}(x) + \mu^2 \sum_x \mathrm{P}(x)
\end{aligned}$$

ここで、第2項の $\sum_x x\mathrm{P}(x)$ は、定義より平均 μ である。また、第3項の $\sum_x \mathrm{P}(x)$ は、確率の総和だから1である。これらを代入して、整理すると、

$$\sigma^2 = \sum_x x^2\mathrm{P}(x) - \mu^2 \tag{3.3}$$

を得る。

上の公式を用いて、サイコロの分散を計算すると、

$$\begin{aligned}\sigma^2 &= \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} - \left(\frac{7}{2}\right)^2 \\ &= \frac{35}{12}\end{aligned}$$

となり、2つの計算結果は確かに一致している。

【例題】　次の2つの場合を比較してみよう。

(1) 100本のくじの中に、賞金100円の当りくじが1本入っている。このくじの中から2本を引く。

(2) 200本のくじの中に、賞金100円の当りくじが2本入っている。このくじの中から2本を引く。

(1) の場合も (2) の場合も、くじを1本引いたときに当る確率は、1/100で同じであることに注意しよう。

(1) の場合

確率変数として獲得賞金をとることにすると、$X = 0,\ 100$ である。2本のくじを引いた場合に100円が当たるのは、1本目が当りか2本目が当りかのいずれかである。

$$\mathrm{P}(100) = \mathrm{P}(1\text{本目が当り}) + \mathrm{P}(2\text{本目が当り})$$

当りくじは1本なので、

$$\begin{aligned}\mathrm{P}(1\text{本目が当り}) &= \mathrm{P}(2\text{本目が外れ}\,|1\text{本目が当り})\mathrm{P}(1\text{本目が当り}) \\ &= \frac{99}{99} = \frac{1}{100}\end{aligned}$$

となる。

$$\begin{aligned}
\mathrm{P}(2\,\text{本目が当り}) &= \mathrm{P}(2\,\text{本目が当り}\,|1\,\text{本目が外れ})\mathrm{P}(1\,\text{本目が外れ}) \\
&= \frac{1}{99}\frac{99}{100} \quad = \frac{1}{100}
\end{aligned}$$

ゆえに、$\mathrm{P}(100) = \dfrac{1}{50}$ を得る。2 本とも外れくじ、すなわち、0 円を獲得する確率は、$\mathrm{P}(0) = 1 - \mathrm{P}(100) = \dfrac{49}{50} = 0.98$ である。ゆえに、賞金の期待値 (この場合は、平均値より期待値の方が適切な気がする) と分散は、

$$\begin{aligned}
\mu &= 0 \times \frac{49}{50} + 100 \times \frac{1}{50} = 2\,\text{円} \\
\sigma^2 &= 0^2 \times \frac{49}{50} + 100^2 \times \frac{1}{50} - 2^2 \\
&= 200 - 4 \\
&= 196\,\text{円}^2
\end{aligned}$$

(2) の場合

(1) と同様に確率変数として獲得賞金をとることにすると、$X = 0,\ 100,\ 200$ となる。それぞれの確率は、

$$\begin{aligned}
\mathrm{P}(0) &= \mathrm{P}(2\,\text{本目が外れ} \cap 1\,\text{本目が外れ}) \\
&= \mathrm{P}(2\,\text{本目が外れ}\,|1\,\text{本目が外れ})\mathrm{P}(1\,\text{本目が外れ}) \\
&= \frac{197}{199} \times \frac{198}{200} \\
&= \frac{19503}{19900}
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(100) &= \mathrm{P}(2\,\text{本目が外れ} \cap 1\,\text{本目が当り}) + \mathrm{P}(2\,\text{本目が当り} \cap 1\,\text{本目が外れ}) \\
&= \mathrm{P}(2\,\text{本目が外れ}\,|1\,\text{本目が当り})\mathrm{P}(1\,\text{本目が当り}) \\
&\quad + \mathrm{P}(2\,\text{本目が当り}\,|1\,\text{本目が外れ})\mathrm{P}(1\,\text{本目が外れ}) \\
&= \frac{198}{199} \times \frac{2}{200} + \frac{2}{199} \times \frac{198}{200} \\
&= \frac{396}{19900}
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(200) &= \mathrm{P}(2\,\text{本目が当り} \cap 1\,\text{本目が当り}) \\
&= \mathrm{P}(2\,\text{本目が当り}\,|1\,\text{本目が当り})\mathrm{P}(1\,\text{本目が当り}) \\
&= \frac{1}{199} \times \frac{2}{200} \\
&= \frac{1}{19900}
\end{aligned}$$

と求まる。

これらは、また、次のようにして求めることもできる。

200 本のくじから 2 本を引く場合の数は、${}_{100}C_2 = 19900$ 通り。2 本とも当りとなる引き方は、1 通り。1 本が当りでもう 1 本が外れとなる引き方は、$2 \times 198 = 396$ 通り。2 本とも外れとなる引き方は、${}_{198}C_2 = 19503$ 通りとなる。獲得賞金が、0, 100, 200 円である確率は、各々、

$$
\begin{aligned}
\mathrm{P}(0) &= \frac{19503}{19900} \approx 0.98005, \\
\mathrm{P}(100) &= \frac{396}{19900} \approx 0.01990, \\
\mathrm{P}(200) &= \frac{1}{19900} \approx 0.00005
\end{aligned}
$$

である。賞金の期待値 (平均値) と分散は、各々、

$$
\begin{aligned}
\mu &= 0 \times \frac{19503}{19900} + 100 \times \frac{396}{19900} + 200 \times \frac{1}{19900} \\
&= \frac{396+2}{199} \\
&= 2\,\text{円} \\
\sigma^2 &= 0^2 \times \frac{19503}{19900} + 100^2 \times \frac{396}{19900} + 200^2 \times \frac{1}{19900} - 2^2 \\
&= 100 \times \frac{(396+4)}{199} - 4 \\
&= \frac{39204}{199} \approx 197.005
\end{aligned}
$$

となる。

(1) と (2) の結果を、表にすると、

	P(0)	P(100)	P(200)	期待値 μ	分散 σ^2
(1) の場合	0.98	0.02	0	2	196
(2) の場合	0.98005	0.01990	0.00005	2	197.005

(1) の場合、確率変数 X の取り得る値は、0 と 100 であった。ところが、(2) の場合には、0, 100, 200 と取り得る値の幅が広がっている。獲得賞金の期待値は同じ 2 円であるが、(1) の場合に較べ、(2) の場合には、P(100) の値が小さくなって、

P(0) と P(200) の値が少し大きくなった。そのため、獲得賞金の額の幅が広がったのに伴い分散の値も大きくなっている。

ちなみに、ここで扱った分布は**超幾何分布**と呼ばれ、サンプリング調査などでも使われる分布である。

連続分布の場合の分散も、平均の場合と同じように、(3.2) 式において、P(x) を $f(x)\,dx$ で、和を積分で置き換えることにより求められる。すなわち、

$$\sigma^2 = \int (x-\mu)^2 f(x)\,dx \tag{3.4}$$

で与えられる。この式は、また、離散分布の場合と同様に変形できて、

$$\sigma^2 = \int x^2 f(x)\,dx - \mu^2 \tag{3.5}$$

と表すことができる。

第4章　離散確率分布

4.1　ベルヌーイ分布

もっとも簡単な確率分布は、「成功か失敗か」のように結果が2つに1つしかない場合である。この分布は、ヤコブ･ベルヌーイにちなんでベルヌーイ分布(Bernoulli distribution)と呼ばれる。確率変数Xを導入し、成功の場合を$X=1$、失敗の場合を$X=0$とし、その各々の確率を、p、$q=1-p$としよう。

$$
\begin{aligned}
P(1) &= p, \\
P(0) &= q = 1-p
\end{aligned}
$$

この分布における確率変数xの平均μは、

$$\mu = 1\times\ p + 0\times\ q = p$$

分散σ^2は、

$$\sigma^2 = (1-p)2 + (0-p)^2 = pq$$

となる。

上記のような試行を独立に何回も続けて行ったものを、**ベルヌーイ試行**(Berunoulli trials)と言う。これから重要な確率分布のいくつかが導かれる。

4.2　幾何分布

「成功か失敗か」のベルヌーイ試行において、1回の試行で成功する確率をp、失敗する確率を$q=1-p$とする。このような試行次々と行う中で、1回から$x-1$回、連続して失敗を繰り返し、x回目にはじめて成功する確率を$\mathrm{P}(x)$とすると、それは、

$$\mathrm{P}(x) = pq^{x-1} \qquad \text{ただし、} x = 1,\ 2,\ 3,\ \cdots$$

で与えられる。このような分布を**幾何分布**と言う。

4.2.1　幾何分布の平均と分散

幾何分布の平均と分散を求めてみよう。

平均

幾何分布の平均は、

$$\mu = \sum_{x=1}^{\infty} xpq^{x-1} = p \sum_{x=1}^{\infty} xq^{x-1}$$

で与えられる。この値を求めたい。そこで、

$$\begin{aligned} S_1(q) &= \sum_{x=1}^{\infty} xq^{x-1} \\ &= 1 + 2q + 3q^2 + 4q^3 + 5q^4 + \cdots \end{aligned} \tag{4.1}$$

とおく。(4.1) 式の両辺に q を掛けて、元の式から引くと、

$$\begin{array}{lrcccccccccccc} & S_1(q) & = & 1 & + & 2q & + & 3q^2 & + & 4q^3 & + & 5q^4 & + & \cdots \\ -) & qS_1(q) & = & & & q & + & 2q^2 & + & 3q^3 & + & 4q^4 & + & \cdots \\ \hline & (1-q)S_1(q) & = & 1 & + & q & + & q^2 & + & q^3 & + & q^4 & + & \cdots \end{array}$$

ここで、

$$S_0(q) = 1 + q + q^2 + q^3 + q^4 + \cdots$$

とおくと、

$$\begin{array}{lrcccccccccccc} & S_0(q) & = & 1 & + & q & + & q^2 & + & q^3 & + & q^4 & + & \cdots \\ -) & qS_0(q) & = & & & q & + & q^2 & + & q^3 & + & q^4 & + & \cdots \\ \hline & (1-q)S_0(q) & = & 1 & & & & & & & & & & \end{array}$$

より、

$$S_0(q) = \frac{1}{1-q} \tag{4.2}$$

したがって、

$$S_1(q) = \frac{S_(q)}{1-q} = \frac{1}{(1-q)^2} = \frac{1}{p^2} \tag{4.3}$$

ここで、$p + q = 1$ を用いた。

ゆえに、

$$\mu = qS_1(q) = \frac{1}{p}$$

を得る。

分散

幾何分布の分散は、

$$\sigma^2 = \sum_{x=1}^{\infty} x^2 pq^{x-1} - \mu^2$$

で与えられる。そこで、

$$\begin{aligned} S_2(q) &= \sum_{x=1}^{\infty} x^2 q^{x-1} \\ &= 1 + 4q + 9q^2 + 16q^3 + 25q^4 + 36q^5 + \cdots \end{aligned} \tag{4.4}$$

が求まれば良い。平均を求めたときと同様に、(4.4) 式に q を掛けて、元の式から引くと、

$$\begin{array}{rrcccccccccccc} & S_2(q) & = & 1 & + & 4q & + & 9q^2 & + & 16q^3 & + & 25q^4 & + & \cdots \\ -) & qS_2(q) & = & & & q & + & 4q^2 & + & 9q^3 & + & 16q^4 & + & \cdots \\ \hline & (1-q)S_2(q) & = & 1 & + & 3q & + & 5q^2 & + & 7q^3 & + & 9q^4 & + & \cdots \end{array}$$

ところが、

$$\begin{aligned} & 1 + 3q + 5q^2 + 7q^3 + 9q^4 + \cdots \\ = \; & 2S_1(q) - S_0(q) \end{aligned}$$

なので、

$$(1-q)S_2(q) = 2S_1(q) - S_0(q) \tag{4.5}$$

ゆえに、

$$S_2(q) = \frac{2S_1(q) - S_0(q)}{1-q}$$

この式に (4.2) 式と (4.3) 式を代入すると、

$$S_2(q) = \frac{2}{(1-q)^3} - \frac{1}{(1-q)^2} = \frac{2}{p^3} - \frac{1}{p^2}$$

ここで、$p = 1 - q$ を用いた。ゆえに、分散は、

$$\begin{aligned}\sigma^2 &= pS_2(q) - \mu^2 \\ &= p\left(\frac{2}{p^3} - \frac{1}{p^2}\right) - \frac{1}{p^2} \\ &= \frac{1}{p^2} - \frac{1}{p} \\ &= \frac{q}{p^2}\end{aligned}$$

を得る。

幾何分布まとめ

確率分布 ： $\mathrm{P}(x) = pq^{x-1},$ ただし、$p + q = 1, \quad x = 1,\ 2,\ 3,\ \cdots$

平均 ： $\mu = \dfrac{1}{p}$

分散 ： $\sigma = \dfrac{q}{p^2}$

【例題】

コインを投げたとき、表または裏が出る確率は、等しく 1/2 であるとする。すなわち、$p = q = 1/2$。コインを表が出るまで投げ続けるとすると、x 回目に初めて表が出る確率は、

$$\mathrm{P}(x) = \frac{1}{2^x}$$

で与えられる。x の平均、すなわち、表が出るまでにコインを放る回数の平均は、

$$\mu = \frac{1}{\frac{1}{2}} = 2$$

で与えられる。

x の分散は、

$$\sigma^2 = \frac{\frac{1}{2}}{\left(\frac{1}{2}\right)^2} = 2$$

である。

4.3　2 項分布

一回の試行で成功する確率を p、失敗する確率を $q = 1 - p$ とする。n 回の試行で、x 回成功する確率は、

$$\mathrm{P}(x) = {}_nC_x p^x q^{n-x}, \ (0 \leqq \ x \leqq \ n)$$

で与えられる。このような分布を **2 項分布**と言い、$B(n, p)$ で表わす。

確率の総和

$$\begin{aligned}\sum_{x=0}^{n} \mathrm{P}(x) &= \sum_{x=0}^{n} {}_nC_x p^x q^{n-x} \\ &= (p+q)^n\end{aligned}$$

ここで、$p + q = 1$ なので、

$$\sum_{x=0}^{n} \mathrm{P}(x) = 1$$

を得る。すなわち、確率の総和は 1 である。

平均と分散

2 項分布の平均は、

$$\begin{aligned}\mu &= \sum_{x=0}^{n} x\mathrm{P}(x) \\ &= \sum_{x=0}^{n} x \, {}_nC_x p^x q^{n-x}\end{aligned}$$

ところが、

$$x_nC_x = x \cdot \frac{n!}{x!(n-x)!} = n \cdot \frac{(n-1)!}{(x-1)!\,((n-1)-(x-1))} = n_{n-1}C_{x-1}$$

だから、

$$\begin{aligned}\mu &= np \sum_{x=1}^{n} {}_{n-1}C_{x-1} p^{x-1} q^{(n-1)-(x-1)} \\ &= np \sum_{x'=0}^{n-1} {}_{n-1}C_{x'} p^{x'} q^{(n-1)-x'} \\ &= np(p+q)^{n-1}\end{aligned}$$

ここで、$p+q=1$ だから、結局、

$$\mu = np \tag{4.6}$$

を得る。この結果は、直観的に理解しやすい。1回で成功する確率が p の試行を n 回繰り返すと、成功する回数の平均は、(試行回数) $\times$ (1回の試行で成功する確率) で与えられる。

分散は、

$$\sigma^2 = \sum_{x=0}^{n} x^2 \mathrm{P}(x) - \mu^2 \tag{4.7}$$

を計算すればよい。第1項、

$$\sum_{x=0}^{n} x^2 \mathrm{P}(x) = \sum_{x=1}^{n} x^2 \frac{n!}{x!(n-x)!} p^x q^{n-x} \tag{4.8}$$

の x^2 を $x(x-1)-x$ と書き直すと、

$$\begin{aligned} \sum_{x=0}^{n} x^2 \mathrm{P}(x) &= \sum_{x=1}^{n} [x(x-1)+x] \frac{n!}{x!(n-x)!} p^x q^{n-x} && (4.9) \\ &= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} && (4.10) \end{aligned}$$

(4.10) 式第1項は、

$$\sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x q^{n-x} = n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x}$$

ここで、$x' = x-2,\ n' = n-2$ とおくと、

$$\begin{aligned} \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} &= \sum_{x'=0}^{n'} \frac{n'!}{x'!(n'-x')!} p^{x'} q^{n'-x'} \\ &= (p+q)^{n'} \\ &= 1 \end{aligned}$$

(4.10) 式第2項は、

$$\begin{aligned} \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x} &= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \\ &= np(p+q)^{n-1} \\ &= np \end{aligned}$$

以上の計算から、

$$\sum_{x=0}^{n} x^2 \mathrm{P}(x) = n(n-1)p^2 + np = n^2p^2 + np(1-p)$$

を得る。また、これを、$\mu = np$ とともに、(4.7) 式に代入すると、

$$\sigma^2 = npq$$

となる。

例題　硬貨を n 回投げたとき、表がでた回数を x とおくと、x 回表が出る確率は、$B(n, 1/2)$ に従う。

1 枚の効果を投げたとき、結果は裏か表かの 2 通りである。3 回投げた場合 ($n = 3$) に面の出方は、全部で $2^3 = 8$ 通り である。簡単のために、表 (heads) を H で、裏 (tails) を T で表すことにして、全ての場合を表にすると、

x	1 回目	2 回目	3 回目
0	T	T	T
1	T	T	H
	T	H	T
	H	T	T
2	T	H	H
	H	T	H
	H	H	T
3	H	H	H

硬貨の表と裏の面の出方は全て同様に確からしく、各々$p = q = \dfrac{1}{2}$ とすると、8 通りの面の出方の確率は、全て $\left(\dfrac{1}{2}\right)^3 = \dfrac{1}{8}$ である。3 回投げて x 回表がでるのは、?3C_x 回あるので、

$$\mathrm{P}(x) = \frac{{}_3C_x}{8}$$

となる。この分布の平均と分散は、

$$\begin{aligned} \mu &= np = \frac{3}{2} \\ \sigma^2 &= npq = \frac{3}{4} \end{aligned}$$

第5章　正規分布

5.1　正規分布と標準正規分布

確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < +\infty$$

これを、$N(\mu,\ \sigma^2)$ で表す。確率変数がすべての値をとる確率は 1 である。これは、

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, xd = 1$$

と表される。

連続分布の場合には、確率変数の値が特定の値 x の場合の確率をいうのは意味をなさない。なぜなら、その確率は常にゼロだからである。連続変数の場合に意味があるのは、確率変数 X が、ある値 a から別の値 b までの範囲にある確率である。これは、

$$P(a < x < b) = \int_a^b f(x)\, dx$$

で与えられる。

変数変換 $z = \dfrac{x-\mu}{\sigma}$ をおこなうと、

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

となる。これは、平均が 0 で分散が 1 だから、$N(0,\ 1)$ と表される。これを標準正規分布と言う。

正規分布 $N(\mu,\ \sigma^2)$ と標準正規分布の関係は、

$$P(a < x < b) = P_{標準}\left(\frac{a-\mu}{\sigma} < z < \frac{a-\mu}{\sigma}\right)$$

で与えられる。

正規分布において、特に重要な確率の値は、平均値からのずれが標準偏差 σ の 1 倍、2 倍 3 倍のどの場合である。

$$\begin{aligned} P(\ -\sigma < x < \ \ \sigma) &= P_{標準}(-1 < z < 1) = 0.683 \\ P(-2\sigma < x < 2\sigma) &= P_{標準}(-1 < z < 1) = 0.955 \\ P(-3\sigma < x < 3\sigma) &= P_{標準}(-1 < z < 1) = 0.997 \end{aligned}$$

5.2 二項分布の正規分布による近似

二項分布

$$\mathrm{P}(x) = {}_nC_x p^x q^{n-x}$$

の平均 μ、分散 σ^2 は、それぞれ、

$$\mu = np, \quad \sigma^2 = npq$$

と表されるが、試行回数 n が十分に大きい場合には、同じ平均と分散の正規分布

$$f(x) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(x-np)^2}{2npq}}$$

によって良く近似される。

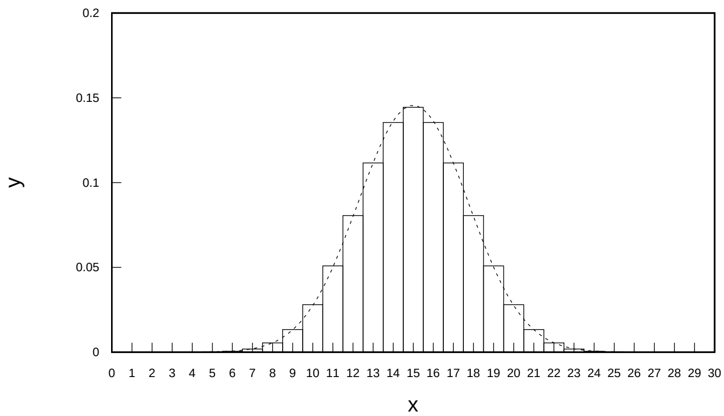


図 5.1: 二項分布と正規分布

図は、$n = 30,\ p = q = 1/2$ の場合の 2 項分布 (棒グラフ) と、同じ平均値と分散の正規分布 (点線) の比較である。両者は、よく一致していることが分かる。

第6章　統計的推論

6.1　検定

一回の試行で成功する確率を p、失敗する確率を q とすると、n 回の試行で x 回成功する確率は、二項分布

$$\mathrm{P}(x) = {}_nC_x p^x q^{n-x}$$

で与えられ、x の平均は $\mu = np$、分散は $\sigma^2 = npq$ で与えられる。

今、n 回の試行をおこなって成功した回数が s 回であったとしよう。n が十分大きい場合に、s の取り得る値の範囲は、おおよそ、

$$\mu - 3\sigma \leqq s \leqq \mu + 3\sigma \tag{6.1}$$

である。つまり、n 回の試行を独立に何度も繰り返した場合、99.7%はこの領域に収まり、この領域からはみ出す確率は僅か 0.3%、つまり 1000 回に 3 回程度である。

ここで、(6.1) 式を、$\hat{p} = s/n$ とおき、更に、μ を np で、σ を $\sqrt{npq}$ で書き換えて、$\hat{p}$ の範囲を求めると、

$$p - 3\sqrt{\frac{pq}{n}} \leqq \hat{p} \leqq p + 3\sqrt{\frac{pq}{n}} \tag{6.2}$$

が得られる。

例として、理想的なサイコロを 100 回振って 6 の目が出る回数 x の確率分布を調べてみよう。(6.1) 式で、$n = 100,\ p = 1/6$ とおくと、$np \approx 16.7,\ 3\sqrt{npq} \approx 11.2$ だから、

$$5.5 \leqq s \leqq 27.8 \qquad \text{または、} \qquad 0.055 \leqq \hat{p} \leqq 0.278$$

を得る。そして、

$$P(6 \leqq s \leqq 27) = \sum_{x=6}^{27} \mathrm{P}(x) = 0.9965$$

である。

もしも、100振って6の目が出た回数が、5回以下だったり28回以上だった場合には、マトモなサイコロとは考えにくい。もちろん、この判断が間違っている可能性もあるが、その確率は、わずか0.35%である。これを、

『このサイコロが正常だという仮説は、**有意水準**0.35%で棄却できる。』

と言う。

それでは、100回振って6の目が27回出た場合には、何が分かるであろうか。この場合には、『正常ではないとは言えない』、または、『異常とは言えない』ことが分かるだけである。決して、正常であることが分かったわけでは無い。

図6.1に正常なサイコロを100回振った時に6の目が出る回数の確率分布を示す。比較のために、平均と分散の等しい正規分布を点線で描いた。

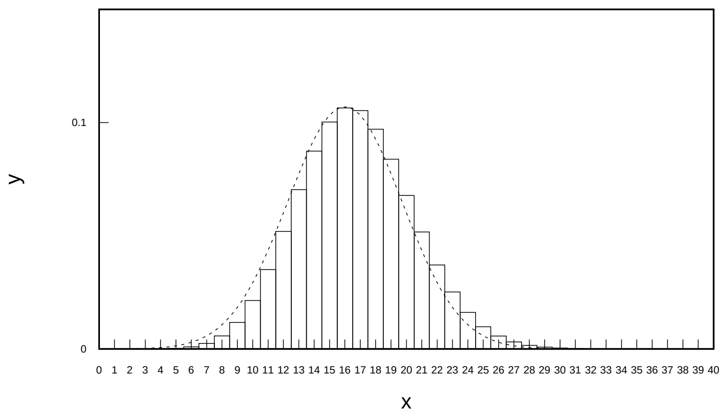


図 6.1: 正常なサイコロを100回振った時に6の目が出る回数の確率分布

サイコロ100回振った場合には、$\hat{p} = 0.06$や$\hat{p} = 0.27$であっても、それが異常とは判断できなかった。サイコロが正常かどうかもっと精度良く判定するためには、振る回数を増やせばよい。振る回数を10倍の1000回にすると、$\sqrt{\frac{pq}{n}}$は、$1/\sqrt{10} = 0.316$倍になるので、(6.2)式は、

$$0.131 \leqq \hat{p} \leqq 0.202$$

となる。

6.2 推定

今度は、観測結果から $\hat{p}$ の値が得られたときに、p の値の範囲を推定する場合について考えよう。(6.2) 式を、p について解くと、

$$\hat{p}-3\sqrt{\frac{pq}{n}} \leqq p \leqq \hat{p}+3\sqrt{\frac{pq}{n}} \tag{6.3}$$

両端の式は、いまだ未知の p と q を含んでいる。これらを値を観測値 $\hat{p}$ と $\hat{q}$ で置き換えても大きな違いはないであろう。そうすると、

$$\hat{p}-3\sqrt{\frac{\hat{p}\hat{q}}{n}} \leqq p \leqq \hat{p}+3\sqrt{\frac{\hat{p}\hat{q}}{n}} \tag{6.4}$$

を得る。この区間に入る確率は、99.7%である。そこで、この区間を、**信頼度 (信頼係数)**99.7%の**信頼区間**と呼ぶ。

例として、次のような場合を考えてみよう。ある調査会社が、3600 人を無作為に抽出して現内閣支持率を調査したところ、52%が内閣を支持すると答えた。国民全体での支持率は、どの範囲にあるとみてよいだろうか。

(6.4) 式に、$n=3600,\ \hat{p}=0.52$ を代入すると、$3\sqrt{\dfrac{pq}{n}}=0.025$ だから、

$$0.495 \leqq p \leqq 0.545$$

となる。すなわち、実際の値は、観測値 $\hat{p}$ の $\pm 2.5\%$ の範囲にあると結論づけられる。およそ 1 億人の意識が、僅か 3600 人を調べるだけで分かるのである。これが、さまざまな場面で世論調査が行われる理由である。

世論調査を行う場合に、標本を偏り無く選ぶことが重要である。すなわち、標本の分布が全体の分布に等しい必要がある。これを端的に示したのが、1936 年のアメリカ大統領選挙であった。この年は、現職の大統領で再選を目指す民主党のフランクリン・ルーズベルトと共和党のアルフレッド・ランドンの間で争われた。

総合週刊誌であるリテラリー・ダイジェストは、電話帳や自動車保有者名簿に名前の載っている人を対象に 1000 万枚の調査票を送り 237 万 6523 人の回答を得て、ランドンの勝利を予測した。一方、当時は無名であったジョージ・ギャラップは、独自の調査を行い、結果を新聞各紙に発表した。その内容は、リテラリー・ダイジェストの結果は誤りで、ルーズベルトが 56%の得票で勝つというものであった。ふたを開けてみると、ルーズベルが勝利し、ギャラップ調査会社は一躍有名になった。

1936 年当時、電話帳や自動車保有者名簿に名前の載るのは、比較的裕福な人々で、共和党支持者が多かった。ところが、リテラシー・ダイジェストの調査から漏れた低所得者層の人々には民主党支持者が多かったのである。

There are three kinds of lies: lies, damned lies, and statistics.

Mark Twain