

1 事象と確率

1.1 母集団・標本・事象・試行

確率や統計では、母集団 (population)、標本 (sample)、事象 (event)、試行 (trial) などと言う言葉が使われる。あまり難しいことは考えずに、例を挙げておく。大阪大学に在籍する学生の喫煙率を調査したいとする。全員にアンケートすれば”全数調査”であって、そこに確率や統計の出番はない。全数調査が難しいときは、ランダムに選ばれた学生に対してアンケートを行い、その結果から全学生の喫煙率を推定することになる。この場合、母集団 Ω は全学生、試行はランダムサンプリング (無作為抽出) でアンケート調査をすること、標本は選ばれた学生、事象は選ばれた学生が喫煙者か非喫煙者かというアンケート結果である。

では、”ダイスを転がす”実験の場合はどうか。調べたいのはたとえば1の目が出る確率であるでしょう。この問題設定はきちんと言うと「ダイスを無限回投げたとすれば、1の目が出る確率はいくらか」であって、「無限回投げる」ことが暗黙に仮定されている。ダイスを投げるのが試行であり、それは「無限回投げる」中からランダムサンプリングされた標本である。母集団 Ω は無限回の試行から成る”仮想的”なものであることに注意。事象はもちろん各試行でどの目が出たかである。

なお、統計解析は標本の”無作為抽出”ができていたとの前提で行なうが、実はこの無作為抽出をどのように実現するかが難しい場合が多い。これは「実験計画法」の問題である。

1.2 確率測度

ダイスを転がして特定の目が出る”確率”などは、直感的にもわかりやすいが、世の中には直感的には理解しにくい確率もある。どのような場合に”確率”を考えてよいかは数学的に定義されている。難しく考えないことにして、以下の三つの条件を満たす P を確率測度とよぶ。

1. 事象 A に対して、 A の確率 $P(A)$ は0と1のあいだにある

$$0 \leq P(A) \leq 1 \quad (1)$$

2. 全事象のいずれかは必ず実現するので、”全事象の確率” P は1。どの事象も実現しない確率、いいかえると”空事象の確率” P は0。ダイスを転がしたとき、1から6までのどれかの目が出る確率は1、1から6までのどの目も出ない確率は0である。

3. 互いに素な事象 A と B に対して

$$P(A \cup B) = P(A) + P(B) \quad (2)$$

ふたつ以上の事象についても同様にこれが成り立つ。

ただし $A \cup B$ は事象 A と B を合せた事象、互いに素とは共通部分をもたないことである。ダイスでいうと、 A が1の目、 B が2の目が出るという事象だとすれば、 $A \cup B$ は”1の目か2の目のいずれかが出る”という事象である。また、”1の目が出る”と”2の目が出る”は

互いに素だが、“1か2のいずれかが出る”と“2か3のいずれかが出る”は互いに素ではない。あるいは、選ばれた学生が“女性である”と“男性である”は互いに素だが、“女性である”と“喫煙者である”は互いに素ではなく、“女性喫煙者”が共通部分である。

1.3 条件つき確率

事象 A と B が確率的に独立であるとは

$$P(A \cap B) = P(A)P(B) \quad (3)$$

が成り立つことをいう。 $A \cap B$ はふたつの事象の共通部分を意味する。

さて、 A が実現するという条件下での B の確率を考えるような場合がある。たとえば、“選ばれた学生が女性であったときにその学生が喫煙者である確率”など。そのような確率を条件つき確率とよび、 $P(B|A)$ で表す。

あきらかに、“ A と B がともに実現する”確率は、“ A が実現し、その条件下で B が実現する確率”に等しいはずだから、

$$P(A \cap B) = P(B|A)P(A) \quad (4)$$

したがって

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (5)$$

である。ただし、条件つき確率は実現する順番を考えているとは限らない。アンケートの集計などは全部のデータが集まってから行なうわけで、全データの中から“女子学生の喫煙率はどのくらいかな”などと条件をつけて事象を選んでいくことになる。その場合は、全く同じデータから逆に“喫煙者の中での女性の比率”を考えることもできる。つまり $P(A|B)$ と $P(B|A)$ の両方を考えてもよい。 $P(A \cap B)$ はもちろん実現する順番とは無関係なので

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B) \quad (6)$$

がなりたつ。

さて、 A と B が独立なら

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B) \quad (7)$$

つまり、 A が実現しようがしまいが、 B の実現確率はその影響を受けないということで、当然の結果である。

全事象の集合 Ω が互いに素である事象に分けられるとき、つまり

$$\Omega = A_1 + A_2 + \cdots + A_n \quad (8)$$

となるときは、

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i) \quad (9)$$

である。これは A_i のうちのいずれかは必ず実現することから明らかだろう。たとえば、学生の喫煙率は”学生が男子であってかつ喫煙者である率”と”学生が女子であってかつ喫煙者である率”を足したものである。

1.4 ベイズの法則

上のように全事象の集合 Ω が互いに素である事象に分けられて

$$\Omega = A_1 + A_2 + \cdots + A_n \quad (10)$$

であるとき

$$P(B|A_i)P(A_i) = P(A_i|B)P(B) \quad (11)$$

より

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad (12)$$

が成り立つ。これをベイズの法則あるいはベイズの規則とよぶ。ここで $P(A_i)$ は B が実現するかどうかと無関係なので事前確率、 $P(A_i|B)$ は B が実現したという条件下でのものなので事後確率とよばれる。

このベイズの規則は、観測結果から「原因の確率」を求めるのに使うことができる。

例：（「数理統計学」（稲垣宣生）より）

それぞれ 10 個の玉がはいったふたつの箱がある。箱 1 にはいっている玉の色と個数は赤 5、白 3、黒 2 である。一方、箱 2 にはいっている玉は赤 2、白 3、黒 5 である。箱 1 からランダムに選んだ一個の玉を箱 2 に移し、そののちに箱 2 から一個の玉を取り出したところ、黒だったとする。この場合、箱 1 から 2 に移された玉がやはり黒であった確率を求めたい。

求めたいのは、箱 2 から黒が取り出されたという条件下で、箱 1 から箱 2 に黒玉が移された確率である。実際に行なった行為とは順番が逆であることに注意。行為と逆順の条件つき確率を考えてもいっこうにかまわない。

ベイズの法則を使えば

$$P(\text{移黒} | \text{出黒}) = \frac{P(\text{出黒} | \text{移黒})P(\text{移黒})}{P(\text{出黒} | \text{移黒})P(\text{移黒}) + P(\text{出黒} | \text{移赤})P(\text{移赤}) + P(\text{出黒} | \text{移白})P(\text{移白})} \quad (13)$$

各項の意味はわかるだろう。

これらに数を代入する

$$P(\text{移黒}) = \frac{2}{10}, P(\text{移赤}) = \frac{5}{10}, P(\text{移白}) = \frac{3}{10} \quad (14)$$

$$P(\text{出黒} | \text{移黒}) = \frac{6}{11}, P(\text{出黒} | \text{移赤}) = \frac{5}{11}, P(\text{出黒} | \text{移白}) = \frac{5}{11} \quad (15)$$

であるから

$$P(\text{移黒} | \text{出黒}) = \frac{3}{13} \quad (16)$$

となる。箱 2 から黒玉が取り出されたという観測事実によって、黒玉が移されていた確率は観測前の $\frac{2}{10} = 2.6/13$ よりも $0.4/13$ だけ高まったことになる。もっとも、この結果をさらにどう解釈するかは、また別の問題である。

例：ある病気の検査は、その病気にかかっているならば 99% の確率で正しく陽性 (+) と診断するが、逆に病気にかかっていない健康な人でも 1% の確率で誤って陽性と診断してしまうことがわかっている。また、別の疫学調査によって、その病気は 10 万人にひとりがかかっているものとしよう (0.001%)。この検査で陽性と診断された人が本当に病気にかかっている確率はどれくらいだろうか。

これは、陽性と診断された原因が”本当に病気にかかっているせい”なのか”健康なのに検査の誤差のせい”なのかの確率を求めようというものである。ベイズの法則を使えば

$$P(\text{病} | +) = \frac{P(+ | \text{病})P(\text{病})}{P(+ | \text{病})P(\text{病}) + P(+ | \text{健})P(\text{健})} \quad (17)$$

である。各項の意味はわかるだろう。

数値を代入する。

$$P(\text{病}) = 0.00001, P(\text{健}) = 1 - P(\text{病}) = 0.99999 \quad (18)$$

$$P(+ | \text{病}) = 0.99, P(+ | \text{健}) = 0.01 \quad (19)$$

であるから

$$P(\text{病} | +) = \frac{0.99 \times 0.00001}{0.99 \times 0.00001 + 0.01 \times 0.99999} \simeq 0.001 \quad (20)$$

つまり、陽性と診断されても、本当に病気である確率は 0.1% に過ぎない。この結果は結局、「あなたがもし、ものすごく珍しい病気かもしれない、と言われたとしても、たいていの場合それは間違いである」という常識を数値で裏づけたものと言える。非常にまれな病気の診断精度を上げたければ、直感的には病人を正しく陽性と診断する確率を上げればよいように思えるが、実は逆で、健康な人を誤って陽性と診断する確率を下げなくてはならないのである。

これはほかの科学実験の場合にも当てはまる。教訓は「大発見をしたと思ったときはよく考え直せ。たいていの場合、それは勘違いだ」ということになるうか。

2 確率分布

2.1 確率変数・分布関数・密度関数

確率変数とは、値の起こりやすさに確率が与えられるような変数で、母集団の特性をあらわすものである。たとえば、大阪大学の学生を全数調査して身長分布を調べておけば、ランダムに選んだひとりの学生の身長がある値をとる確率というものは決まる。別にあらかじめ全数調査しなくても、身長の値の起こりやすさに確率が与えられるはずであることはわかる。そういう量が確率変数である。また、喫煙者なら 1、非喫煙者なら 0 となるような変数を設定することもできる。これも値が喫煙率で決まる確率変数である。また、コインを 10 個同時に投げたときに表の出る個数などは典型的な確率変数である。

そういった確率変数 X に対して、その分布関数 $F(x)$ というものを考える。

$$F(x) = P(X \leq x) \quad (21)$$

これは、「変数 X の値が x 以下であるような確率」を意味する。この分布関数を使えば、たとえば X の値が a 以上 b 以下である確率は

$$P(a \leq X \leq b) = F(b) - F(a) \quad (22)$$

と F の差であらわされる。

ダイスの目のように整数値しかとらないような離散変数の場合は、先に X が値 x_i をとる確率 $f_i = P(X = x_i)$ を考えておいて

$$F(x) = \sum_{x_i \leq x} f_i \quad (23)$$

で分布関数が求められる。和は $x_i \leq x$ であるようなすべての x_i について行なう。もっとも、これでは分布関数を考えるご利益はわからない。

X が特定の値をとる確率ではなく、特定の値以下である確率を考えるのは、確率変数 X が連続値の場合も扱えるようにするためである。たとえば、0 から 1 までの数直線上にランダムに点を打つことを考えよう。この場合、数直線上の幅 l の領域に点がかかる確率は l であるから、それがたとえば「ぴったり 0.1」になる確率は 0 である（点は幅が 0）。しかし、「0.1 以下」である確率は 0.1 となる。この場合、ある値になる確率を考えるのはナンセンスで、どれほど小さくても幅を持った領域を考えなくてはならない。身長や体重なども厳密に言えば連続値をとる量なので、適当な範囲を設定しないかぎり「ある値をとる」確率はいつでも 0 である。我々が現実に身長や体重を語る際には、身長なら 1mm 程度、体重なら 100g 程度の幅を暗黙のうちに仮定しているのではないだろうか。

しかし、常に”範囲”を指定しなくてはならないのはめんどくさい。そこで登場するのが、密度関数 $f(x)$ である。これは $F(x)$ の微分

$$f(x) = \frac{dF(x)}{dx} \quad (24)$$

で定義される。つまり、これは分布関数 $F(x)$ のグラフの傾きである。 $f(x)$ の意味を考えよう。微小な幅 Δx を与えたとき、 $x \leq X \leq x + \Delta x$ である確率は

$$P(x \leq X \leq x + \Delta x) = F(x + \Delta x) - F(x) \quad (25)$$

である。 Δx が充分小さければ、 x から $x + \Delta x$ の微小範囲で $F(x)$ のグラフは直線とみなせるだろう。その傾きは $f(x)$ である。すると、

$$F(x + \Delta x) = F(x) + f(x)\Delta x \quad (26)$$

と書ける。したがって、

$$P(x \leq X \leq x + \Delta x) = f(x)\Delta x \quad (27)$$

である。つまり、微小範囲については $f(x)$ にその幅をかけたものが、その範囲内の値が実現する確率となる。幅をかけて確率になるので、いわば $f(x)$ は”確率の密度”であり、そのために密度関数と呼ばれる。

範囲が微小でない場合はどうか。微分の逆操作が積分であるから、 $a \leq X \leq b$ である確率は

$$F(b) - F(a) = \int_a^b f(x)dx \quad (28)$$

によって求められる。あるいは、

$$F(x) = \int_{-\infty}^x f(x)dx \quad (29)$$

もちろん、 $f(x)$ は

$$0 \leq f(x) \leq 1 \quad (30)$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad (31)$$

という性質を持つ。ちなみに、上の式は $F(x)$ が単調増加であるという自明な性質から。

2.2 期待値 (平均値)・モーメント・分散

確率変数 X の期待値 (平均値とよぶことも多い) $E(X)$ は

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx \text{ (連続値)} \quad (32)$$

$$E(X) = \sum_i x_i f_i \text{ (離散値)} \quad (33)$$

で定義される。また、 X の任意の関数 $h(X)$ の期待値は

$$E\{h(X)\} = \int_{-\infty}^{\infty} h(x)f(x)dx \text{ (連続値)} \quad (34)$$

$$E\{h(X)\} = \sum_i h(x_i)f_i dx \text{ (離散値)} \quad (35)$$

で定義される。以下では面倒なので、連続値の場合だけを書くことにする。特に $h(X) = X^k$ (k は自然数) の場合を X の k 次のモーメントとよぶ。

$$E\{X^k\} = \int_{-\infty}^{\infty} x^k f(x)dx \quad (36)$$

X の 1 次モーメントは X の期待値そのものである。

X の分散 $V(X)$ は 2 次モーメントと 1 次モーメントの 2 乗の差で定義され

$$V(X) = E(X^2) - \{E(X)\}^2 \quad (37)$$

これが、”期待値との差の 2 乗の期待値” に等しいことはすぐに確かめられる。

$$E\{(X - E(X))^2\} = E\{X^2 - 2XE(X) + (E(X))^2\} = E(X^2) - 2E\{XE(X)\} + E\{E(X)^2\} \quad (38)$$

ここで、 $E(X)$ は確率変数ではないただの値なので、 $E(X)$ の期待値は単に $E(X)$ であるから、

$$E\{(X - E(X))^2\} = E(X^2) - 2E(X)E(X) + \{E(X)\}^2 = E(X^2) - \{E(X)\}^2 \quad (39)$$

である。

分散は期待値からのずれの目安であるが、2 乗なので平均値とは単位が違う。ずれの目安としては、この平方根である標準偏差 S がよく用いられる

$$S = \sqrt{V(X)} \quad (40)$$

2.3 モーメント母関数・特性関数・キユムラント

モーメント母関数 (Moment generating function) は

$$M(t) = E\{e^{tX}\} = \int_{-\infty}^{\infty} e^{tx} f(x)dx \quad (41)$$

で定義される。これがモーメントの母関数と呼ばれる理由は、

$$\left. \frac{d^k M(t)}{dt^k} \right|_{t=0} = \int_{-\infty}^{\infty} \left. \frac{d^k}{dt^k} e^{tx} \right|_{t=0} f(x)dx = \int_{-\infty}^{\infty} x^k f(x)dx = E(X^k) \quad (42)$$

と、 $M(t)$ を k 回微分することによって k 次モーメントが求められるからである。ただし、 $\left. \right|_{t=0}$ という記号は、 t による微分を行なったのちに $t = 0$ を代入することを意味している。あるいは、同じことだが、指数関数のテーラー展開

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (43)$$

を使って

$$M(t) = E \left\{ \sum_{k=0}^{\infty} \frac{(tX)^k}{k!} \right\} = \sum_{k=0}^{\infty} \frac{1}{k!} E(X^k) = 1 + E(X) + \frac{1}{2}E(X^2) + \frac{1}{3!}E(X^3) + \dots \quad (44)$$

と書ける。

また、似た量として特性関数 (characteristic function) があり、

$$\Psi(t) = E\{e^{itX}\} = \int_{-\infty}^{\infty} e^{itx} f(x) dx \quad (45)$$

で定義される。 i は虚数単位である ($i^2 = -1$)。これは要するに” X をフーリエ変換したもの”である。モーメント母関数と同様

$$\left. \frac{d^k \Psi(t)}{dt^k} \right|_{t=0} = i^k E(X^k) \quad (46)$$

が成り立つ。実は、分布関数によってはモーメント母関数が存在しない (無限大になる) ことがあるのに対し、フーリエ変換である特性関数はいつでも存在する。モーメント母関数が存在すれば、特性関数とのあいだに

$$\Psi(t) = M(it) \quad (47)$$

$$\Psi(it) = M(-t) \quad (48)$$

の関係がある。

特性関数 (およびモーメント母関数) の重要な性質が一致性である。分布関数が等しければ特性関数が等しいのは当然だが、逆に特性関数が等しければ分布関数も等しい。また、特性関数は無限次までのすべてのモーメントを含むから、特性関数が等しいというのはつまりあらゆるモーメントが等しいということである。

X の k 次のキュムラント (cumulant) C_k は次の式で定義される

$$G(t) \equiv \log M(t) = \sum_{k=1}^{\infty} \frac{1}{k!} C_k \quad (49)$$

和は $k = 1$ から始まることに注意。 $G(t)$ はキュムラント母関数とよばれる。左辺に $M(t)$ の展開式を代入して、さらに対数関数のテーラー展開

$$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots \quad (50)$$

を使い、両辺で t が同じ次数となる項の係数を等しいとおけば

$$C_1 = E(X) \quad (51)$$

$$C_2 = E(X^2) - \{E(X)\}^2 = V(X) \quad (52)$$

などの関係が求められる。一般に k 次のキュムラントは k 次以下のモーメントの組み合わせであらわせる。とりあえずは、1 次のキュムラントが期待値、2 次のキュムラントが分散であるということだけ頭にはいっていただろう。

3 代表的な分布関数

3.1 離散一様分布 $DU(n)$

X は n 通りの値 x_1, x_2, \dots, x_n をとり、その実現確率はすべて等しく

$$f_i = P(X = x_i) = \frac{1}{n} \quad (53)$$

で与えられる。コインの裏表やダイスの目なら、分布はそれぞれ $DU(2)$ と $DU(6)$ である。

$X = 1, 2, 3, \dots, n$ なら

$$E(X) = \frac{1}{n} \sum_{k=1}^n k = \frac{n+1}{2} \quad (54)$$

$$E\{X^2\} = \frac{1}{n} \sum_{k=1}^n k^2 = \frac{(n+1)(2n+1)}{6} \quad (55)$$

$$V(X) = E\{X^2\} - \{E(X)\}^2 = \frac{n^2 - 1}{12} \quad (56)$$

となる

3.2 二項分布 $B_N(n, p)$

確率 p で成功し、 $1 - p$ で失敗するような試行 (ベルヌイ試行とよぶ) を独立に n 回行ったとき、その成功回数を確率変数 X とする。 $X = x$ (失敗回数は $n - x$) である確率は

$$f(x) = {}_n C_x p^x (1 - p)^{n-x} \quad (57)$$

ただし、 ${}_n C_x$ は n 個の中から x 個を選ぶ組み合わせの数で

$${}_n C_x = \frac{n!}{x!(n-x)!} \quad (58)$$

である。このような確率分布 $f(x)$ を二項分布という。名前は二項展開

$$(a + b)^n = \sum_{m=0}^n {}_n C_m a^m b^{n-m} \quad (59)$$

に由来する。つまり、 $1 = \{p + (1 - p)\}^n$ (これは任意の n でなりたつ) を二項展開すると

$$1 = \{p + (1 - p)\}^n = \sum_{x=0}^n {}_n C_x p^x (1 - p)^{n-x} = \sum_{x=0}^n f(x) \quad (60)$$

となる。

二項分布の平均と分散は、二項展開を使って求めることができる

$$E(X) = \sum_{x=0}^n x_n C_x p^x (1-p)^{n-x} = n \sum_{x=1}^n {}_{n-1}C_{x-1} p^x (1-p)^{n-x} \quad (61)$$

$$= np \sum_{x=0}^{n-1} {}_{n-1}C_x p^x (1-p)^{n-x-1} = np \{p + (1-p)\}^{n-1} = np \quad (62)$$

また、少々技巧的だが $E(X^2)$ ではなく $E\{X(X-1)\}$ を求めると

$$E\{X(X-1)\} = \sum_{x=0}^n x(x-1) {}_n C_x p^x (1-p)^{n-x} = n(n-1) \sum_{x=2}^n {}_{n-2}C_{x-2} p^x (1-p)^{n-x} \quad (63)$$

$$= n(n-1)p^2 \sum_{x=0}^{n-2} {}_{n-2}C_x p^x (1-p)^{n-x-2} = n(n-1)p^2 \{p + (1-p)\}^{n-1} = n(n-1)p^2 \quad (64)$$

したがって

$$V(X) = E(X^2) - \{E(X)\}^2 = E\{X(X-1)\} + E(X) - \{E(X)\}^2 = np(1-p) \quad (65)$$

例：裏表の出かたに偏りのないコインを n 回投げたとき、表が x 回出る確率は

$$f(x) = {}_n C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} \quad (66)$$

表の出る回数の期待値は

$$E(X) = np = \frac{n}{2} \quad (67)$$

これは当然の結果である。標準偏差は

$$\sqrt{V(X)} = \sqrt{np(1-p)} = \frac{\sqrt{n}}{2} \quad (68)$$

たとえば 10 回投げるとき、期待値は 5 回、標準偏差は 1.6 回程度である。この意味はまたあとで。

3.3 ポアソン分布 $P_0(\lambda)$

ポアソン分布は、まれにしか起きない事象を大量に観測したときに、事象が起きる回数の確率分布である。たとえば、単位時間あたりの放射性元素の崩壊数は、一個の原子が崩壊する確率は低くても原子数が多い (1 モルなら 10^{23} 個) ことから、多数の崩壊が観測され、その回数はポアソン分布に従う。また、「数理統計学」(稲垣宣生)によれば、一日あたりの交通事故の数もポアソン分布に従うらしい。一台の車が事故を起こす確率は極めて低くても、台数が多いので事故は毎日起きており、まれな事象の大量観測の例になっている。

ポアソン分布に従う確率変数 X の確率分布は一個のパラメータ λ によって

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (69)$$

と定義される。

ポアソン分布の期待値と分散を求めるには、特性関数を使うのが簡単。特性関数の定義から

$$\Psi(t) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} e^{itx} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{1}{x!} (\lambda e^{it})^x = \exp \{ \lambda (e^{it} - 1) \} \quad (70)$$

これを用いて

$$E(X) = \frac{1}{i} \frac{d\Psi(t)}{dt} \Big|_{t=0} = \lambda \quad (71)$$

また、 $V(X)$ は 2 次のキュムラントであるから

$$V(X) = - \frac{d^2 \log \Psi(t)}{dt^2} \Big|_{t=0} = \lambda \quad (72)$$

3.4 連続変数の一様分布 $U(\alpha, \beta)$

確率変数 X は連続変数で、 $\alpha \leq X \leq \beta$ の区間で一様分布するとき、密度関数は

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & (\alpha \leq x \leq \beta) \\ 0 & (\text{それ以外の範囲}) \end{cases} \quad (73)$$

期待値と分散はそれぞれ

$$E(X) = \int_{\alpha}^{\beta} x f(x) dx = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x dx = \frac{\alpha + \beta}{2} \quad (74)$$

また

$$E(X^2) = \int_{\alpha}^{\beta} x^2 f(x) dx = \frac{1}{\beta - \alpha} \frac{\beta^3 - \alpha^3}{3} \quad (75)$$

より

$$V(X) = \frac{(\beta - \alpha)^2}{12} \quad (76)$$

4 正規分布

4.1 分布関数

正規分布 $N(\mu, \sigma^2)$ の密度関数はふたつのパラメータ μ と σ^2 をもちいて

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (77)$$

で定義される。

4.2 標準正規分布

$N(0, 1)$ は標準正規分布とよばれる。この密度関数を特に $\phi(x)$ と書くことにすると

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (78)$$

である。一般の正規分布 $N(\mu, \sigma^2)$ の密度関数は z 変換

$$Z = \frac{X - \mu}{\sigma} \quad (79)$$

によって、標準正規分布の密度関数 $\phi(z)$ に変換される。したがって、標準正規分布の性質がわかっていれば、一般の正規分布の性質もわかる。なお、係数の分母から σ が姿を消したのは、積分変数の変換からくるものであるが、単に全領域での積分が

$$\int_{-\infty}^{\infty} \phi(x) dx = 1 \quad (80)$$

となるように決めたとってもよい。

ここで、ガウス積分

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi} \quad (81)$$

である。これを計算するもっとも簡単な方法は

$$I = \int_{-\infty}^{\infty} e^{-x^2/2} dx = \int_{-\infty}^{\infty} e^{-y^2/2} dy \quad (82)$$

とおいて

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right) \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy \right) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy e^{-(x^2+y^2)/2} \quad (83)$$

ここで xy 座標の作る平面を考えると、 $x^2 + y^2$ は座標原点から点 (x, y) までの距離の 2 乗なので

$$r^2 = x^2 + y^2 \quad (84)$$

とする。一方、積分

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \quad (85)$$

は平面の全面積を積分せよという意味だが、別のやりかたでもできる。原点を中心とする半径 r と $r + \Delta r$ のふたつの円を考えると、 Δr が十分に小さければこのふたつの同心円に挟まれた領域の面積は $2\pi r \Delta r$ である。平面の全面積について積分するには、これを r の全範囲について積分すればよい。つまり

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy = \int_0^{\infty} 2\pi r dr \quad (86)$$

これより

$$I^2 = \int_0^{\infty} 2\pi r e^{-r^2/2} dr \quad (87)$$

ところが

$$\frac{d}{dr} e^{-r^2/2} = -r e^{-r^2/2} \quad (88)$$

なので

$$I^2 = -2\pi \int_0^{\infty} \left(\frac{d}{dr} e^{-r^2/2} \right) dr \quad (89)$$

しかし、微分して積分すれば元に戻るだけなので、結局

$$I^2 = -2\pi \left[e^{-r^2/2} \right]_0^{\infty} = 2\pi \quad (90)$$

が得られる

4.3 期待値・分散

標準正規分布の期待値は

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx = 0 \quad (91)$$

である。これは、 $x e^{-x^2/2}$ が 0 を中心として反対称（正側と負側で絶対値が等しくて逆符号）なので、正側の積分と負側の積分が打ち消すからである。

分散を求めるには少々トリッキーだが、以下の積分を考える

$$J(\lambda) = \int_{-\infty}^{\infty} e^{-\lambda x^2/2} dx \quad (92)$$

なぜなら、これを用いて

$$\frac{dJ(\lambda)}{d\lambda} \Big|_{\lambda=0} = \int_{-\infty}^{\infty} \left(\frac{d}{d\lambda} e^{-\lambda x^2/2} \right) \Big|_{\lambda=0} dx = - \int_{-\infty}^{\infty} \left(\frac{x^2}{2} e^{-\lambda x^2/2} \right) \Big|_{\lambda=0} dx = - \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \quad (93)$$

となるからである。また、

$$\sqrt{\frac{\lambda}{2\pi}} e^{-\lambda x^2/2} dx \quad (94)$$

は $N(0, 1/\sqrt{\lambda})$ の分布関数にほかならないから、全領域で積分して 1。したがって

$$J(\lambda) = \sqrt{\frac{2\pi}{\lambda}} \quad (95)$$

これより

$$E(X^2) = -2 \frac{1}{\sqrt{2\pi}} \frac{dJ(\lambda)}{d\lambda} \Big|_{\lambda=0} = -2 \frac{1}{\sqrt{2\pi}} \left(\frac{d}{d\lambda} \sqrt{\frac{2\pi}{\lambda}} \right) \Big|_{\lambda=0} = 1 \quad (96)$$

したがって

$$V(X) = E(X^2) - \{E(X)\}^2 = 1 \quad (97)$$

つまり、標準正規分布は期待値 0、分散 1 の正規分布である。

一般の正規分布 $N(\mu, \sigma^2)$ の期待値と分散は Z 変換の逆変換

$$X = \mu + \sigma Z \quad (98)$$

を考えればよい。 Z は標準正規分布に従うから

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu \quad (99)$$

また、

$$V(X) = V(\mu + \sigma Z) = E\{(\mu + \sigma Z)^2\} - \{E(\mu + \sigma Z)\}^2 = \sigma^2 \quad (100)$$

つまり、 μ と σ^2 はそれぞれ正規分布の期待値と分散なのだった (ということは、 σ が標準偏差)。したがって、任意の正規分布は期待値と分散のふたつの値で特徴づけられる。

4.4 モーメント母関数・キュムラント

一般の正規分布のモーメント母関数も求めておく。

$$M(t) = E\{e^{tX}\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{tx - (x-\mu)^2/2\sigma^2} dx = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \quad (101)$$

最後の結果は、指数の肩を平方完成すれば導ける。

これより、キュムラントの母関数は

$$G(t) = \log M(t) = \mu t + \frac{\sigma^2 t^2}{2} \quad (102)$$

これが t の 2 次までしか含まないことから、正規分布では 3 次以上のキュムラントが全て 0 であることがわかる。一致性から、逆に 3 次以上のキュムラントが全て 0 であるような確率分布は正規分布に限られることもわかる。

4.5 正規分布の値

標準正規分布に従う確率変数 Z が、 $0 \leq Z \leq z$ の範囲にある確率は標準正規分布表に与えられている。もっとも、今なら統計解析ソフトがあるので、紙に書かれた表を使う機会はあまりないかもしれない。一般の正規分布については、正規分布から Z 変換の逆変換によって求めることができる。

代表的な値を挙げておく。

$$P(-\sigma \leq X - \mu \leq \sigma) \simeq 0.682$$

$$P(-2\sigma \leq X - \mu \leq 2\sigma) \simeq 0.954$$

$$P(-3\sigma \leq X - \mu \leq 3\sigma) \simeq 0.997$$

この値を見れば、標準偏差の2倍あるいは3倍程度が分布の幅の目安として使われる理由が納得できるだろう。

5 2変数の同時確率分布

5.1 同時分布・周辺分布

各サンプルのもつふたつの性質をあらわす確率変数 X と Y の同時分布を考える。たとえば、学生の身長と体重の分布を調べる場合などがこれにあたる。以下、 X と Y はともに連続変数としておくが、離散変数の場合も同様に考えることができる。

同時分布関数は

$$F(x, y) = P(X \leq x, Y \leq y) \quad (103)$$

で定義される。 X または Y だけの分布関数は周辺分布関数とよばれる

$$F_1(x) = P(X \leq x) = F(x, \infty) \quad (104)$$

$$F_2(y) = P(Y \leq y) = F(\infty, y) \quad (105)$$

1変数の場合を拡張して、同時密度関数は

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) \quad (106)$$

と定義される。 XY 平面中に $\Delta x \times \Delta y$ の微小長方形を考えると、その範囲の値が実現する確率が

$$f(x, y) \Delta x \Delta y \quad (107)$$

だと思っておけばよい。同時密度関数は

$$f(x, y) \geq 0 \quad (108)$$

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy f(x, y) = 1 \quad (109)$$

という性質をもつ。

X または Y だけの密度関数は周辺密度関数とよばれる。

$$f_1(x) = \frac{dF_1(x)}{dx} = \int_{-\infty}^{\infty} f(x, y) dy \quad (110)$$

$$f_2(y) = \frac{dF_2(y)}{dy} = \int_{-\infty}^{\infty} f(x, y) dx \quad (111)$$

変数 X と Y が統計的に独立であれば

$$f(x, y) = f_1(x) f_2(y) \quad (112)$$

がなりたつ。

5.2 共分散・相関係数

周辺分布については、一変数と同じなので、期待値や分散を考えることができる。値をそれぞれ

$$E(X) = \mu_1, E(Y) = \mu_2 \quad (113)$$

$$V(X) = \sigma_1^2, V(Y) = \sigma_2^2 \quad (114)$$

としておく。

2変数関数 $h(X, Y)$ の期待値は

$$E\{h(X, Y)\} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy h(x, y) f(x, y) \quad (115)$$

で定義される。もし、 $h(X, Y)$ が X だけの関数と Y だけの関数の直積型

$$h(X, Y) = h_1(X)h_2(Y) \quad (116)$$

で、かつ X と Y が統計的に独立であれば

$$E\{h(X, Y)\} = E\{h_1(X)\}E\{h_2(Y)\} \quad (117)$$

がなりたつ。直積型であっても、 X と Y が独立でなければこれはなりたないことに注意

2変数の関係をあらわす重要な量が共分散である。これは名前の通り、分散を2変数に拡張したもので、以下で定義される

$$Cov(X, Y) = E\{(X - \mu_1)(Y - \mu_2)\} = E(XY) - E(X)E(Y) = E(XY) - \mu_1\mu_2 \quad (118)$$

X と Y を Z 変換した量

$$Z_1 = \frac{X - \mu_1}{\sigma_1}, Z_2 = \frac{Y - \mu_2}{\sigma_2} \quad (119)$$

の共分散は相関係数 ρ とよばれる。定義から

$$\rho \equiv Corr(X, Y) \equiv Cov(Z_1, Z_2) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} \quad (120)$$

相関係数は

$$-1 \leq \rho \leq 1 \quad (121)$$

であるが、 $\rho = \pm 1$ となるのは

$$Z_2 = \pm Z_1 \quad (122)$$

あるいは書き直して

$$Y = \mu_2 \pm \frac{\sigma_2}{\sigma_1}(X - \mu_1) \quad (123)$$

という関係があるときに限られる。つまり、 Y と X が線形関係にある場合である。

また、 X と Y が統計的に独立であれば $E(XY) = E(X)E(Y)$ なので、共分散は 0、したがって相関係数も 0 である。ただし、逆は必ずしもなりたない。つまり、一般には相関係数が 0 であっても統計的に独立とはいえない。

相関係数は 2 変数間での線形関係の強さの目安となる。 $\rho > 0$ なら、 X が平均より正にずれているときは Y も正にずれている確率が高く、逆に $\rho < 0$ なら、 X と Y の平均値からのずれは逆符号である確率が高いことを示す。ただし、線形ではない相関関係は相関係数に正しく反映されない。また、相関があることと因果関係とは全く別の問題であるので、結果の解釈には注意が必要である。

5.3 2 変数の和

確率変数 X と Y の定数倍の和

$$U = aX + bY \quad (124)$$

もまた確率変数である。この期待値と分散は

$$E(U) = E(aX + bY) = aE(X) + bE(Y) = a\mu_1 + b\mu_2 \quad (125)$$

$$V(U) = E\{(aX + bY)^2\} - \{E(aX + bY)\}^2 = a^2V(X) + b^2V(Y) + 2ab \times Cov(X, Y) \quad (126)$$

U の分散に X と Y の共分散が現れる。

5.4 2 次元正規分布

2 次元正規分布 $N_2(\mu, \Sigma)$ は、期待値ベクトル

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad (127)$$

と、分散共分散行列

$$\Sigma = \begin{pmatrix} V(X) & Cov(X, Y) \\ Cov(X, Y) & V(Y) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad (128)$$

で指定され、分布関数は次式で定義される。

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \frac{(x-\mu_1)^2}{\sigma_1^2} - 2\frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right\} \right] \quad (129)$$

また、 X と Y をそれぞれ z 変換すれば

$$f(z_1, z_2) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp \left[-\frac{1}{2(1-\rho^2)} (z_1^2 - 2\rho z_1 z_2 + z_2^2) \right] \quad (130)$$

である。

$\rho = 0$ の場合は

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left[-\frac{1}{2} \left(\frac{x-\mu_1}{\sigma_1} \right)^2 \right] \times \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left[-\frac{1}{2} \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] = f_1(x)f_2(y) \quad (131)$$

と独立な正規分布関数の積になる。つまり、2次元正規分布の場合は、相関係数が0ならふたつの変数は統計的に独立である。この結論は明らかに2変数以上が正規分布している場合にもなりたつ。

この式を $x + y$ について整理してみると

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_1}{\sigma_1}\right)^2\right] \times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{y - \mu_2}{\sigma_2}\right)^2\right] = f_1(x)f_2(y) \quad (132)$$

$f(z_1, z_2)$ をもう少し考えてみる。新しいふたつの変数

$$u = z_1 + z_2, v = z_1 - z_2 \quad (133)$$

を用いると

$$\frac{1}{2(1 - \rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2) = \frac{1}{2}\left(\frac{u^2}{1 + \rho} + \frac{v^2}{1 - \rho}\right) \quad (134)$$

と書き換えられる。つまり、新しい確率変数 $U = Z_1 + Z_2$ と $V = Z_1 - Z_2$ はそれぞれ独立な正規分布に従う。2次元正規分布は、このように必ず独立な正規分布の積に書き直せる。

もともとの Z_1 と Z_2 が統計的に独立 ($\rho = 0$) な場合について、 U の周辺密度関数は

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad (135)$$

つまり、正規分布にしたがうふたつの変数の和は、やはり正規分布に従う。

6 大数の法則と中心極限定理

6.1 標本平均

同じ測定を独立に多数回繰り返すことを念頭において、以下の状況を考える。確率変数 X_1, X_2, \dots, X_n は同一の確率分布に従う独立な確率変数であるとする。すべて、分布の期待値は μ 、分散は σ^2 である。

この変数の (単純) 平均

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \quad (136)$$

を標本平均とよび、やはり確率変数である。 n は標本数。

標本平均の期待値は明らかに

$$E(\bar{X}_n) = \mu \quad (137)$$

また分散は

$$V(\bar{X}_n) = \frac{1}{n^2} [E\{(X_1 + \dots + X_n)^2\} - \{E(X_1 + \dots + X_n)\}^2] = \frac{\sigma^2}{n} \quad (138)$$

つまり、標準偏差は個々の標準偏差の $1/\sqrt{n}$ になる。

6.2 大数の法則 (law of large numbers)

標本数 n を無限大とする極限で、標本平均 \bar{X}_n は母平均 μ に収束する

これが大数の法則である。これは \bar{X}_n の期待値が μ であることとは違う。 $E(\bar{X}_n) = \mu$ は、 \bar{X}_n の値が確率に従って分布し、その分布の期待値が母平均に一致することを述べている。一方、大数の法則は $n \rightarrow \infty$ の極限では \bar{X}_n の値が μ になる確率が 1、つまりかならず (数学的にはいろいろ難しい問題はあるが、それは気にしないことにして) μ に一致することを述べている。

実用上は、測定回数を増やせばその単純平均が母集団の期待値にいくらでも近づく、ということである。

6.3 中心極限定理 (central limit theorem)

標本数 n を無限大とする極限で、 $\sqrt{n}\bar{X}_n$ の分布は正規分布 $N(\mu, \sigma^2)$ に収束する

分散が存在しない (分散が無限大) 分布は例外で、たとえばコーシー分布などがそれにあたる。それらの例外を除き、あらゆる確率分布の標本平均 (の \sqrt{n} 倍) の分布は標本数 $n \rightarrow \infty$ で正規分布に収束する。

実用上は、分散が存在する確率分布の標本平均は、標本数が十分に多いとき、正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従うと思ってよい。

例:

n 個のコインを同時に投げることを考える。確率変数としては、 i 番目のコインが表なら 1 となり裏なら 0 となる変数 $X_i (i = 1, 2, \dots, n)$ を考える。表が出た個数 M は

$$M = X_1 + X_2 + \dots + X_n \quad (139)$$

M は二項分布 $B_N(n, 1/2)$ に従う確率変数である。これは上で述べた標本平均の n 倍であるから、それを考慮すると、 n が大きければ、 M の分布は正規分布 $N(n/2, n/4)$ とみなすことができる。

7 統計的推定

7.1 標本平均・標本分散・不偏標本分散

母集団から n 回の独立なランダムサンプリングを行なう。母集団は平均 μ 、分散 σ_2 をもつが、それらの値は未知である。このようにサンプリングされた標本に関する計算値から、母集団の平均や期待値を推定するのがこの節の目的である。

各回の標本 X_1, X_2, \dots, X_n は同一の確率分布に従う確率変数である。標本平均 (標本の単純

平均)

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \quad (140)$$

も確率変数であり、中心極限定理によれば、 n が大きいときは正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従う。一方、標本分散は

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \quad (141)$$

で定義される。前節の $V(\bar{X})$ とは違う量なので、注意。また、これは「標本平均との差の 2 乗平均」であって「母平均との差の 2 乗平均」ではないことにも注意。標本平均は確率変数である。

標本平均の期待値と分散は前節で見た通り

$$E(\bar{X}) = \mu, V(\bar{X}) = E\{\bar{X}^2\} - \{E(\bar{X})\}^2 = \frac{\sigma^2}{n} \quad (142)$$

であるから、

$$E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2 \quad (143)$$

一方、各 X_i の期待値と分散はそれぞれ μ, σ^2 であるから

$$E(X_i^2) = \sigma^2 + \mu^2 \quad (144)$$

これらを用いると、標本分散の期待値として

$$E(S_X^2) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \sigma^2 - \frac{\sigma^2}{n} \quad (145)$$

つまり、標本分散の期待値は母分散よりも $\frac{\sigma^2}{n}$ だけ小さい。こうなる理由は、母分散との差ではなく、確率変数である標本平均との差を計算しているためである。このような系統的な差を偏差 (bias) とよぶ。この偏差を補正したものが不偏 (unbiased) 標本分散で、

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (146)$$

で定義され、この期待値は母分散と一致する。

これらに対して、実際に観測をおこなえば観測値が x_1, x_2, \dots, x_n などと得られる。観測値の標本平均 (言葉は上と同じだが) は

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) \quad (147)$$

で計算される。これは n が大きいときは正規分布 $N(\mu, \frac{\sigma^2}{n})$ からランダムサンプリングされた値とみなせる。また、観測値の標本分散は同様に

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (148)$$

これは「観測値の標本平均との差の2乗平均」であることに注意。上と同様に観測値の不偏標本分散は、偏差を補正した

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (149)$$

で求められる。

これらより、母平均の推定値としては標本平均 \bar{x} をまた母分散の推定値としては標本不偏分散 $\hat{\sigma}_x^2$ を採用すればよい。

7.2 区間推定

母平均の推定値は \bar{x} でよいとして、それはどの程度信頼できる値だろうか。推定値と真の母平均が完全に一致することはまずありえないから、真の母平均は \bar{x} を中心とする適当な範囲にあると考えられる。その範囲を推定するのが区間推定である。

区間推定の前に次の例を考えてみる

例

裏表の出方が均等なコインを 100 回投げるとき、表が出る回数 X はどの範囲にあると考えられるだろうか。

二項分布 $B_N(100, 1/2)$ は正規分布 $N(50, 25)$ とみなせる。 z 変換すれば、

$$Z = \frac{X - 50}{\sqrt{25}} = \frac{X}{5} - 10 \quad (150)$$

Z は $N(0, 1)$ に従う。この分布では $-3 \leq Z \leq 3$ の範囲に含まれる確率が 99% 以上 (99.7%) であるから、 Z の測定値はほぼ間違いなくこの範囲にはいるはずである。 z 変換の逆変換をおこなうと、この範囲は

$$-3 \leq \frac{X}{5} - 10 \leq 3 \quad (151)$$

より

$$35 \leq X \leq 65 \quad (152)$$

したがって、表の回数はほとんど確実に 35 回から 65 回の間になるはずである。逆にいうと、この範囲をはずれた場合はなにか不自然なことが起きていると疑われる。

この例では、「ほとんど確実」とは 99.7% のことだった。これでは厳しすぎると考えるなら、 $-2 \leq Z \leq 2$ の範囲とすれば、95.4% である。この場合、表の回数は 40 回から 60 回の間となる。

この例を念頭において、表の出た個数を観測することによって表の出る確率を区間推定することを考える。必ずしも裏表の出方が均等かどうかわかっていないコインを 100 回投げる。表の出る確率を p とすれば、表の出る回数 X は二項分布 $B_N(100, p)$ に従う。これは正規分布

$N(100p, 100p(1-p))$ とみなせる。Z 変換

$$Z = \frac{X - 100p}{\sqrt{100p(1-p)}} \quad (153)$$

すれば、Z は $N(0, 1)$ に従う。Z の観測値は 99.7% の確率で $-3 \leq Z \leq 3$ の範囲にはいるから、Z 変換の逆変換により

$$100p - 3 \times 10\sqrt{p(1-p)} \leq X \leq 100p + 3 \times 10\sqrt{p(1-p)} \quad (154)$$

この不等式は X と $100p$ を移項してみればわかるように

$$\frac{X}{100} - 0.3\sqrt{p(1-p)} \leq p \leq \frac{X}{100} + 0.3\sqrt{p(1-p)} \quad (155)$$

と書き換えられる。この不等式は p の範囲を与えているが、両辺に p が現れているので、このままでは p の範囲を推定できない。しかし、 p そのものの推定値は $\frac{X}{100}$ であるから、左辺と右辺の p を推定値でおきかえてしまうことにより

$$\frac{X}{100} - 0.3\sqrt{\frac{X}{100} \left(1 - \frac{X}{100}\right)} \leq p \leq \frac{X}{100} + 0.3\sqrt{\frac{X}{100} \left(1 - \frac{X}{100}\right)} \quad (156)$$

と p の範囲が X だけであらわせたことになる。これが p を区間推定する式である。今の場合、区間は正規分布の 99.7% の範囲なので、99.7% の信頼区間といわれる。これを 99% というのもあると思う。95.4% の信頼区間でよければ、範囲を $\frac{2}{3}$ 倍すればよい。それを単に 95% の信頼区間ということもあると思う。99% の信頼区間といえば、正しくは標準正規分布表から読み取って $-2.575 \leq Z \leq 2.575$ 、また 95% なら、正しくは $-1.96 \leq Z \leq 1.96$ である。

たとえば、100 回投げて 40 回表が出たとすれば、99.7% の信頼区間で推定される p の値はだいたい $0.25 \leq p \leq 0.55$ 、また 95.4% の信頼区間では $0.3 \leq p \leq 0.5$ である。いずれにしても、完全なコインの確率である $p = 0.5$ は信頼区間に (かなり端だが) 含まれる。

ここで、99% 信頼区間とは、真の値がその区間に 99% の確率で含まれるという意味ではなく、観測値を 99% 区間を含むような確率分布 (正規分布) の平均値の範囲を示している。つまり、上のようにして 99% 信頼区間で区間推定された範囲の p であれば、どの p をとっても $N(100p, 100p(1-p))$ の 99% 区間内に観測値 (上の例では 40 回) が含まれる。

7.3 正規分布の平均の区間推定 (母分散が既知の場合)

ここでは正規分布の平均を区間推定することを考えるが、中心極限定理があるので、実用上は多くの場合がこれに含まれる。

母分散 σ^2 が既知であるとする。 n 回の測定の標本平均 (今度は n で割ったもの) \bar{X} を Z 変換すると

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \quad (157)$$

で、これは $N(0, 1)$ にしたがう。95.4% の信頼区間で母平均 μ を区間推定したければ

$$-2 \leq Z \leq 2 \quad (158)$$

したがって

$$\bar{X} - 2 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 2 \frac{\sigma}{\sqrt{n}} \quad (159)$$

である。

一般に、 $N(0, 1)$ に従う確率変数 Z が $-z_\alpha^* \leq Z \leq z_\alpha^*$ に含まれる確率が $1 - \alpha$ であるような z_α^* を標準正規分布の両側 α 点とよぶ。たとえば $z_{0.05}^* = 1.96$ である。これを用いて母平均 μ の $1 - \alpha$ 信頼区間とは

$$\bar{X} - z_\alpha^* \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_\alpha^* \frac{\sigma}{\sqrt{n}} \quad (160)$$

をいう。

7.4 正規分布の平均の区間推定 (母分散が未知の場合)

母分散 σ^2 が未知の場合は、真の母分散のかわりに観測値からの不偏推定値 $\hat{\sigma}_n^2$ を用いることになる。 σ^2 のかわりに $\hat{\sigma}_n^2$ を用いて Z 変換と同様の変換をすることを T 変換とよぶ。

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}_n} \quad (161)$$

これが $N(0, 1)$ に従うなら話は前節と同じなのだが、実はそうではない。 T は自由度 $n - 1$ の t 分布という分布に従うことが知られている。分布が標本数 n に依存することが前節とは大きく違う点であることに注意。 t 分布の値は t 分布表に与えられているので、その値を使って前節と同様に区間推定をすればよい。上と同様に自由度 $n - 1$ の t 分布の両側 α 点を $t_{n-1, \alpha}^*$ とすれば、 μ の $1 - \alpha$ 信頼区間は

$$\bar{X} - t_{n-1, \alpha}^* \frac{\hat{\sigma}_n}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1, \alpha}^* \frac{\hat{\sigma}_n}{\sqrt{n}} \quad (162)$$

と求められる。

7.5 χ^2 分布

Z_1, Z_2, \dots, Z_n が、それぞれが標準正規分布 $N(0, 1)$ に従う独立な n 個の確率変数であるとき、その 2 乗和

$$S^2 = Z_1^2 + Z_2^2 + \dots + Z_n^2 \quad (163)$$

は、自由度 n の χ^2 分布 χ_n^2 に従う。

$n = 1$ のときは簡単で、標準正規分布 $N(0, 1)$ に従う確率変数 Z の 2 乗 $S = Z^2$ について

$$P(0 \leq S \leq s) = P(|Z| \leq \sqrt{s}) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{s}} e^{-x^2/2} dx \quad (164)$$

ここで変数変換

$$y = x^2 \quad (165)$$

を行なうと

$$\frac{dy}{dx} = 2x = 2\sqrt{y} \quad (166)$$

より

$$\frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{s}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_0^s \frac{1}{\sqrt{y}} e^{-y/2} dy \quad (167)$$

が得られる。従って、 S の密度関数は

$$f(s) = \frac{dP(0 \leq S \leq s)}{ds} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2} \quad (168)$$

8 仮説検定

なにか確認したい仮説があって、それが成り立つか成り立たないかを統計的に判断するのが「仮説検定」である。仮説にもいろいろありうるが、ここでは数式で表現できる仮説を扱う。たとえば、「喫煙者は非喫煙者に比べて肺がんのリスクが高い」などは典型的な仮説である。仮説検定にはいろいろと思想的な闘争の歴史があり、ある意味「不都合な真実」にもあふれている。しかし、まずは仮説検定の実例を見てみよう。

とりあえず、ここでは簡単な問題として、前節でも取り上げたコイン投げを考えてみよう。たとえば、コインを 100 回投げたところ表が 60 回で裏が 40 回出たとする。普通は、表の出る確率も裏の出る確率も等しく、このときにはたまたま表がたくさん出ただけと考えるだろう。しかし、実はこのコインは表と裏の重さのバランスが少しずれていて、本当に表が出やすい可能性も考えられる。そこで、「このコインは表が出やすい」という仮説を立てることにする。

仮説検定の特徴は証明したい仮説に対する帰無仮説 (null hypothesis) というものを考えるところにある。帰無仮説というのは、おおざっぱに言うと「特に変わったことは起きていない」という仮説である。今の例では「コインの表と裏は等確率で出る」が帰無仮説となる。それに対して、検証したい仮説のほうは対立仮説と呼ぶならわしである。検証したい方が対立とは釈然としないかもしれないが、しきたりなのでしかたない。そして、帰無仮説が成り立っていると仮定しても観測結果が説明できるかどうかを考える。つまり、裏表が等確率に出るコインを 100 回投げたとき、表裏 60:40 というのはどれくらい珍しい事態なのかを考える。もし、これがとても珍しいのなら、帰無仮説は間違いだったと考えていいだろう。これを「帰無仮説が棄却された」と言う。

ところで、帰無仮説が棄却されたら、対立仮説が立証されたことになるのだろうか。それは対立仮説の種類にもよるが、一般には帰無仮説が棄却されたからといって対立仮説が正しいとは限らない。しかし、帰無仮説が棄却されたのだから、とりあえず対立仮説が正しかったことにしておこうと考える場合が多い。これは本来は論理的帰結ではないので、思想的対立の元となっている。

今の実例に即して計算してみよう。帰無仮説に従えばコインの表と裏の数は確率 $1/2$ の二項分布 $B(100, 1/2)$ に従う。したがって、100 個投げて表が n 回出る確率は単純に

$$P(n) = \frac{1}{2^{100}} {}^{100}C_n \quad (169)$$

である。電卓で計算してみると表が 60 回出る確率は約 1% である。1 度の試行で出るにはいささか珍しいような気がする。しかし、ここで知りたいのは表がぴったり 60 回出る確率というよりは、表が 60 回以上出る確率だろう。つまり

$$p(n \geq 60) = \frac{1}{2^{100}} \sum_{n=60}^{100} {}^{100}C_n \quad (170)$$

が知りたいのではないかと思う。この計算をそのままやってもいいが、ちょっとやっかいなので、前節と同様に二項分布はほぼ正規分布で近似できることを使おう。二項分布 $B_N(100, 1/2)$ は正規分布 $N(50, 25)$ とみなせるので、 Z 変換して、

$$Z = \frac{X - 50}{\sqrt{25}} = \frac{X}{5} - 10 \quad (171)$$

が $N(0, 1)$ に従う。このとき、 $X = 60$ は $Z = 2$ なので、標準正規分布で $P(Z \geq 2)$ を求めると

$$P(Z \geq 2) = 0.023 \quad (172)$$

であることがわかる。どのようにして求めたかという、前節で $-2 \leq Z \leq 2$ となる確率は 0.954 だった。欲しいのは $Z \geq 2$ の確率であり、それは

$$P(Z \geq 2) = 1 - P(Z \leq 0) - P(0 \leq Z \leq 2) = 1 - 0.5 - P(-2 \leq Z \leq 2)/2 \quad (173)$$

から求められる。つまり、コインを 100 個投げたとき表が 60 個以上出る確率は 2.3% である。これはどういう意味だろうか。コインを 100 個投げるという試行を 1000 回行えば、そのうちの 23 回程度の試行では表が 60 個以上出るだろうということである。このように、帰無仮説のもとでそのような事象が起きる確率を p 値と呼ぶ。今の例では p 値は 2.3% である。さて、その程度に珍しいことが 1 度の試行で起きたとき、どう解釈するか。

ひとつの可能性は、珍しいとは言っても所詮は 100 回試せば 2 回起きる程度の珍しさなのだから、単に珍しいことが起きたでいいではないかというもの。もうひとつは、100 回試してようやく 2 回起きる程度に珍しいことが起こったのだから、このコインは表と裏が等確率で出るとは言えないのではないかというものだ。そのどちらであるかを判定する絶対的な方法はない。

そこで、「帰無仮説に従うなら 100 回に一回しか起きない珍しいできごとが起きたのだから、帰無仮説は間違っていると考えよう」とか「20 回に一回しか起きない珍しいできごとが起きたのだから」とか、判定基準をあらかじめ決めておくのである。観測してから判定基準をあと付けで決めてはならない。あくまでも、観測を行う前に決めておく。本来であれば、これくらいの数の試行をす

るのだから判定基準はこれくらいにしておくのがよからうというところまでを実験の計画に含めるのが正しい。しかし、分野や学術雑誌によってはまったくそういう実験計画とは関係なしに「判定基準はこれこれ」と決められていることがあり、それは科学的に誠実とはとても言えないと思う。

さて、気をとりなおして、たとえば今は 20 回に 1 回くらいしか起きない珍しいことが起きたら帰無仮説が棄却されるものと決めておいたとしよう。つまり帰無仮説に従うと 0.05 以下の確率でしか起こらないできごとが起こったら、帰無仮説は棄却されるものとする。すると、60 個のコインが表になったのは、それよりも珍しいことが起きたのだから、帰無仮説は棄却される。この判定基準を有意水準と呼ぶ。たとえば「有意水準 0.05 で帰無仮説は棄却された」などという。

しかし、それではその有意水準を変えたら棄却されていたものもされなくなったりするのではないかという疑問がわくのは当然である。そして、もちろんそのとおりである。あまり緩い条件にすると、おうおうにして、その問題が生じる。

すぐにわかるように、帰無仮説が本当に正しいかどうか（それはあらかじめわかっているわけではないし、絶対にわからないかもしれないが）におうじて結果は大きく四つにわけられる

1. 帰無仮説は実際に正しく、そして棄却されなかった
2. 帰無仮説は実際は正しいにもかかわらず、棄却された
3. 帰無仮説は実際は誤りであるにもかかわらず、棄却されなかった
4. 帰無仮説は実際は誤りであり、そして棄却された

このうち 1 と 4 は正しく判定されているので問題ない。それに対し 2 と 3 は別の意味で誤りである。2 を第一種の誤り、3 を第二種の誤りとよぶ。

コイン投げ程度なら実害もないだろうが、たとえば病気の検査ではこの違いは大きな問題となる。たとえば、ある検査の数値がいくら以上なら病気と判定するか。この場合、帰無仮説は「病気ではない」である。判定基準を厳しくすれば第二種の誤り（病気の見落とし）が増え、判定基準を緩くすれば第一種の誤り（健康なのに病気と判定）が増える。このふたつがトレードオフの関係にあることはすぐにわかるだろう。この手のトレードオフはたとえば刑事裁判にも見られる。冤罪を減らしたければ、実際に罪を犯した人がある程度見逃す覚悟が必要となる。

検定とは本来確率的にしか決まらないできごとに、白か黒かの線を引く作業である。それ自体、ほんらいはできないことをやっているとも言えるし、あるいは意思決定とはそのようなものだということもできる。その操作には必ずしも科学だけでは決められない何かが含まれていることは認識すべきである。最近、白か黒かを判定せずに p 値のみを表示する流儀も多く見られる。しかし、それはそれで、各自が各自の基準で「p 値がこれこれだから帰無仮説は棄却されたと考えよう」などと判断しているのだろう。

ところで、たくさん試行すれば統計的な結果も真の結果に近づくだらうと考えたいのは人情というものである。そして、大数の法則は、少なくとも平均値についてその考えが正しいことを保証しているし、区間推定の推定幅も小さくなるのだから推定精度も上がる。それなら、検定だって、たくさん試行すればするほど正しくなるに違いないと考えたいところなのだが、どうもこと検定に限ってはそういうわけにはいかないようである。一般に、試行の数を増やすと帰無仮説は棄却され

やすくなることが知られている。そこで近年は p 値ではなく「効果量」という量を表示することも増えているが、ここではこれくらいにしておこう。推定までの話に比べると、検定はいろいろな意味で気持ちの悪い問題をたくさん含んでいることだけ理解してもらいたい。つまり、「エクセルで計算して有意水準 0.05 で帰無仮説棄却」とか言っていれば済む問題だとは思わないこと