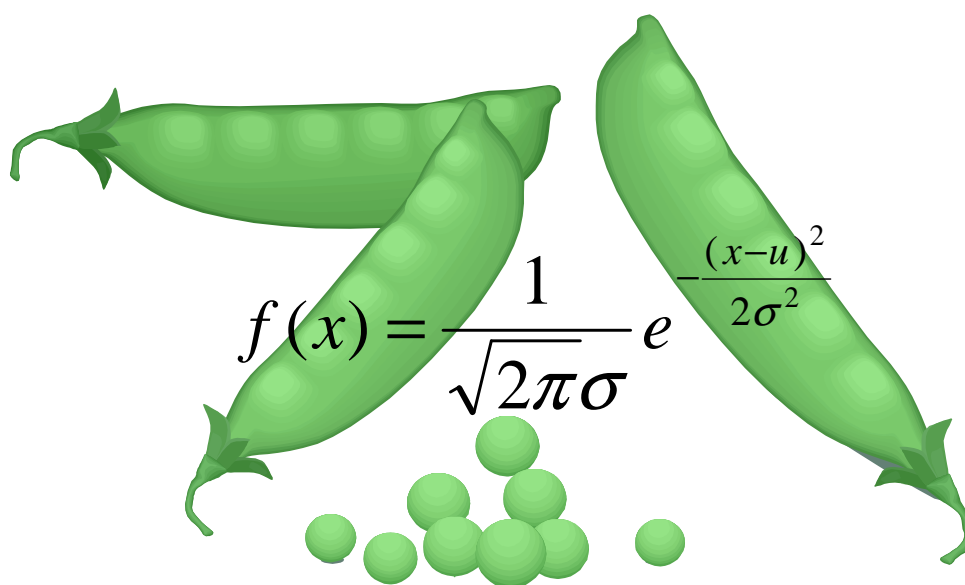


バイオインフォマティクスのための
数理統計学入門*



慶應義塾大学先端生命科学研究所・環境情報学部
専任講師 斎藤輪太郎

平成 22 年 5 月 8 日

*To my wife and Dr. Nozomu Yachie.

序文

統計学は自然科学、社会科学、人文科学などあらゆる分野に応用できる可能性を秘めた学問分野であり、これを習得することには大変大きな意義がある。筆者の専門であるバイオインフォマティクスにとっても、統計学は非常に重要な学問である。

最近では統計処理用言語 R など統計処理ソフトがかなり充実し、原理をそれほど知らなくても複雑な処理を手軽に行えるようになってきた。それとともに、どの統計ソフトを使うと何ができるを幅広く知っていることが重要との考え方が徐々に支配的になりつつある。

確かに重要なのは自然科学や社会科学の中で解決したい問題そのものであって、統計はその手段にしか過ぎない、という考え方もある。

しかしながら、統計の数理的な側面を深く勉強すれば、データの見方が変わってくるし、独自の統計処理を設計できるようになるだろう。そこに待っているのはデータの中に潜む新たな発見かも知れない。

また純粋な興味として、乱雑に見える様々な現象の多くが実は統計学で扱われる正規分布という分布に従う、というだけでも驚くべき事ではないだろうか。いったいどうしてそのようなことが起こるのだろうか。

本書では統計の原理を数学を用いてとことんまで追求することを目指した。統計学の素晴らしさ、面白さを少しでも感じて頂ければ幸いである。

2010 年 3 月 19 日 鶴岡タウンキャンパスにて
著者

目次

| | | |
|-----------|-----------------------------|-----------|
| 1 | 確率分布 | 4 |
| 1.1 | 度数分布 | 4 |
| 1.2 | 離散確率分布 | 7 |
| 1.3 | 連続確率分布 | 8 |
| 1.4 | 確率密度関数の変数変換 | 12 |
| 2 | 期待値 | 13 |
| 3 | 標本分布 | 15 |
| 4 | 二項分布 | 17 |
| 5 | 正規分布 | 20 |
| 6 | ポワソン分布 | 22 |
| 7 | テイラー展開 | 23 |
| 7.1 | 平均値の定理 | 23 |
| 7.2 | テイラーの定理 | 24 |
| 8 | 偏微分 | 25 |
| 9 | 行列式 | 27 |
| 9.1 | 置換 | 27 |
| 9.2 | 行列式の定義 | 30 |
| 9.3 | 行列式の性質 | 31 |
| 9.4 | 内積 | 32 |
| 9.5 | グラム・シュミットの直交化法 | 33 |
| 9.6 | 行列式と超平行面体の超体積との関係 | 34 |
| 10 | 多変量の確率関数 | 35 |
| 10.1 | 多変量の確率関数の定義 | 35 |
| 10.2 | 多変量の確率関数に関する期待値計算 | 38 |
| 10.3 | 多変量の確率関数の変数変換 | 38 |
| 11 | 積率母関数 | 39 |
| 11.1 | 積率母関数の定義 | 39 |
| 11.2 | 確率変数の関数の積率母関数 | 41 |
| 11.3 | 正規分布による二項分布の近似 | 42 |
| 12 | 中心極限定理 | 43 |

1 確率分布

1.1 度数分布

日常生活で我々は極めて多くの「数」に接している。気温、株価、通勤時間、経費、商品の売り上げ高、... と枚挙に暇がない。その中でも経費など我々の意思でかなりコントロールできる数と、気温のように我々の意思ではほとんどコントロールができない数が存在する。商品の売り上げ高などは、一部我々の意思でコントロールできる微妙なものであろう。これから取り上げたいのは、我々の意思ではなかなか完全にコントロールできないような数がどのような法則に従っているかである。

同じ性質を持った多数の数がどのような法則に従っているかを捉える第一歩は度数分布を調べることであろう。これは各値がどれくらいの頻度で観測されるかを表すものである。例えば図 1(a) は、日本人の各年齢別の人口を表したものである。横軸が年齢、縦軸はその年齢の人口を表している。度数分布をこのように縦軸に値、横軸にその値の頻度をとって各棒の位置と高さで各値とその頻度を表したものをヒストグラムと呼ぶ。

この例では、より人口が多い中国やインドで同様のヒストグラムを描くと、ヒストグラム全体の高さも上昇する。人数そのものには関心がなく、日本の人口に対してどの年齢層にどれくらいの割合が分布しているかを知りたいときは、縦軸に人数そのものではなく、該当する年齢層の全体に対する割合をプロットするといいたいだろう。このようにしたのが、図 1(b) である。分布の形状は変わっていないが、縦軸の目盛りが変わっているのが分かるだろう。人口の総数を N 、年齢 i に該当する人口を n_i で表すと、年齢 i に該当する割合 p_i は、

$$p_i = \frac{n_i}{N} \quad (1)$$

となる。ここで、

$$\sum_i p_i = 1 \quad (2)$$

が成立する。このように縦軸を割合にした分布は後ほど述べる確率分布で使われるので、ぜひ慣れて頂きたい。

度数分布に関しては次のような興味深い話がある [4]。第二次大戦後のドイツでは食料が不足し、1 人当たりのパンの配給量は 200g と決められていた。ある学者はパン屋から毎日受け取るパンの重さに疑いを持ち、毎日重さを量り、度数分布を作成した。すると、図 2(a) のように左右がほぼ対称釣り鐘状の分布になったが、中心の位置が 200g より低い位置にあることが分かった。

そこで学者は「君はパンの目方をちゃんと量っていないぞ。」と伝えた。するとパン屋は「パンなんて重いものもあれば、軽いものもありますよ。」そこで学者はヒストグラムを見せ、「本気で 200g にしようとしたならば、グラフは図 2(b) のようになるはずだ。ところが君が作るパンのグラフは 2(a) のようになっている。明らかに目方が不足しているじゃないか。」と告げた。パン屋は「これからは気をつけます。」と言った。

そして次の日から学者に配給されるパンは目方が不足しなくなった。しかし、学者がまた目方のヒストグラムを描いてみたところ、図 2(c) のようなグラフになった。これでパン屋が量を多くした

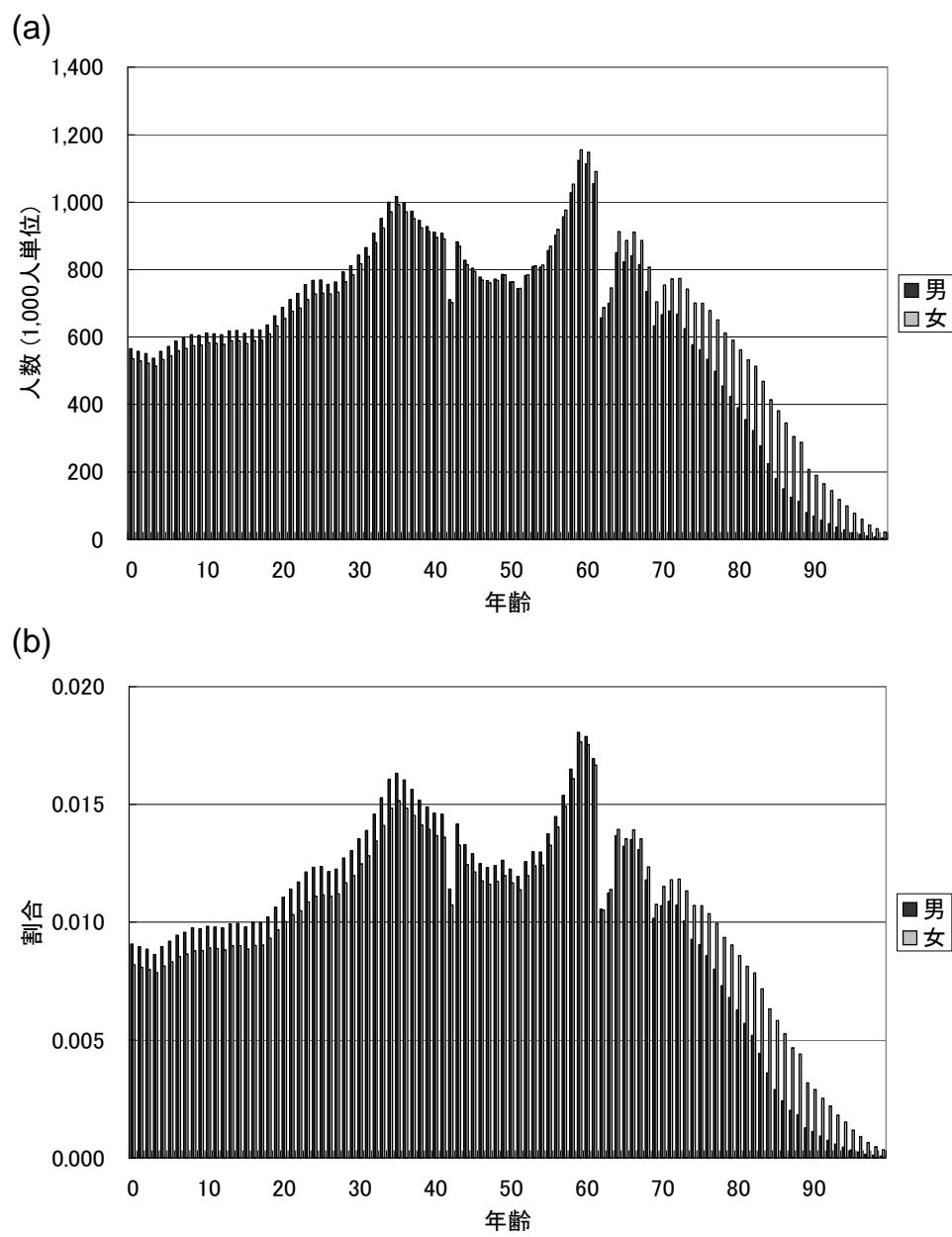


図 1: 2008 年の日本の年齢別人口 (男女別)

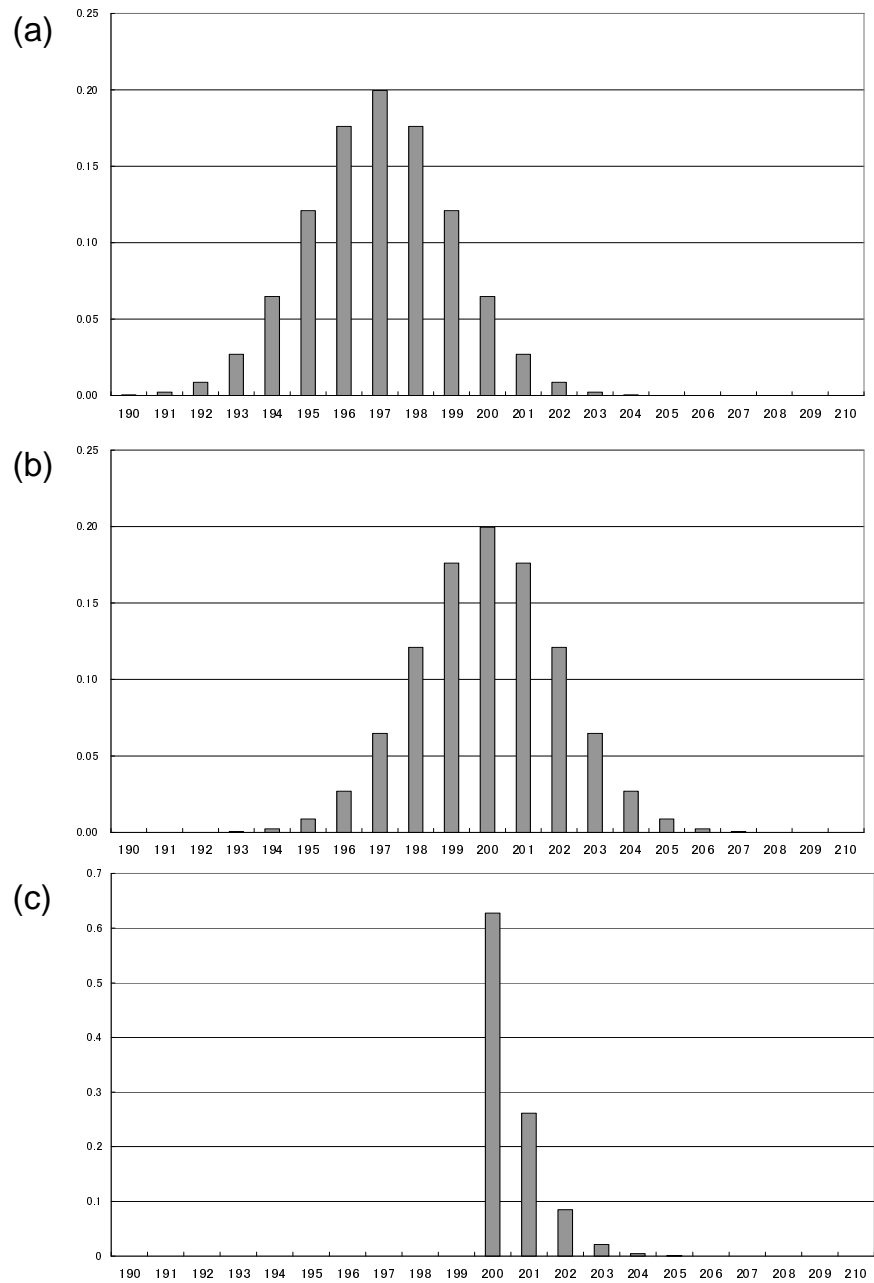


図 2: パン屋から受け取るパンの重量の分布

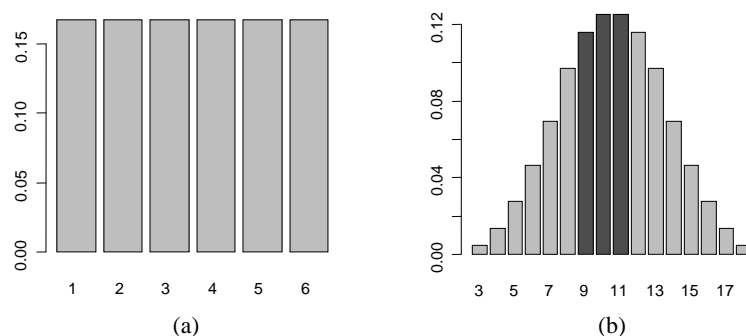


図 3: サイコロを例にした確率分布の例

(a) サイコロの各目が出る確率 (b) サイコロを 3 回振ったときの目の合計とその確率。合計が 9 以上 11 以下の確率に該当する棒の色を濃く示した。

のではなく、学者のところにはだけ目方が足りているパンを選んで渡していることが分かってしまった。度数分布の威力を示す 1 つの例である。

1.2 離散確率分布

変数 X が様々な値を取り、またどのような値を取るかが確率的に決まる時、 X を確率変数と呼ぶ。例えばサイコロを振ったときに出た目を X とすると、 X は確率的に決まるため、確率変数である。 X がある値 x を取る確率を $P(X = x)$ と表す。例えば X をサイコロを振ったときに出る目とすれば、サイコロの目が 2 になる確率は $P(X = 2) = 1/6$ と表される。誤解を招く恐れがないときは、 $P(x)$ と省略する。

X が取り得る値 x_1, x_2, x_3, \dots を横軸に、その値を取る確率 p_1, p_2, p_3, \dots を縦軸にしたヒストグラムを描くと、変数 X の特性が分かる。例えば横軸にサイコロの目、縦軸にその目が出る確率をヒストグラムで表すと図 3(a) のようになるし、またサイコロを 3 回振ったときの目の合計を横軸、その確率を縦軸にすると、図 3(b) のようになる。このように変数 X の確率的な分布を表したものを確率分布と呼ぶ。

図 3 の場合、ヒストグラムの各棒の幅を 1 とすれば、各棒の面積が確率変数が取り得る各値の確率となり、さらに以下の 2 つの重要な性質が得られる。

1. 棒の面積の合計が 1 となる
2. X が a 以上、 b 未満となる確率 $P(a \leq X < b)$ は、以下の式で与えられる。

$$P(a \leq X < b) = \sum_{a \leq x_i < b} P(X = x_i) \quad (3)$$

逆に必要条件として、確率分布は $P(X = x_i) \geq 0$ かつ上記 2 つの性質を満たすもの、と約束する。こうすることによって様々な確率分布に対してその特性の計算方法を統一できる。サイコロを 3 回

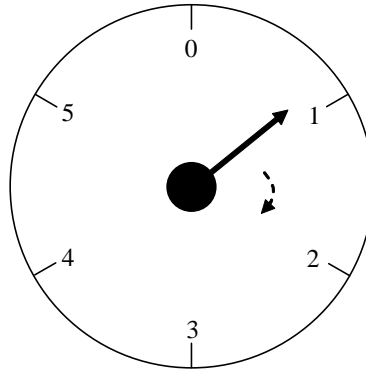


図 4: サイコロルーレット

振る例で、目の合計が 9 以上 11 以下になる確率は $P(9) + P(10) + P(11)$ 、すなわち図 3 の色の濃い棒の面積の和で表されることが分かる。

上記の例では確率分布をヒストグラムで表したが、確率変数 X が値 x を取る確率 $P(x)$ を数式で表現できれば最も客観的である。これを考えると、図 3(a) の分布は $P(x) = \frac{1}{6}$ と表されるし、図 3(b) の分布は $P(x) = \sum_{i=0}^{\lfloor \frac{1}{6}x-2 \rfloor} \frac{3(x-6i-1)(x-6i-2)}{i!(3-i)!} (-1)^i$ である¹。但し、 $\lfloor \alpha \rfloor$ は α 以下の最大の整数を表す (小数以下を切捨てた整数)。このように確率変数 X が各値 x を取る確率を関数で表したものを確率関数と呼ぶ。

確率分布と度数分布の間には密接な関係がある。 N 回の試行で $X = x_i$ となる現象が n_i 回起こったとする。但し、その現象が起こる確率は $P(X = x_i)$ である。このとき、 N が大きくなるにつれて、 n_i/N は $P(X = x_i)$ に近づいてゆく。

問題 1-1: 各年に特定の都市 A で大地震が起きる確率は $\frac{1}{60}$ であるとする。今年の始めから数えて x 年の間に 1 回も地震が起らない確率を $P(x)$ としたとき、 $P(x)$ の確率関数を求めよ。

1.3 連続確率分布

サイコロの例では確率変数 X は離散値、すなわち 1,2,3,4,5,6 と整数値を取り、2.5 など間の数を取ることはなかったし、1 から 6 という範囲の中で取り得る値は 6 個と有限であった。確かに現実で起こる確率的と思われる現象の中には、気象を例に取ればある地域における年間の台風発生数、落雷数など離散的な数で表されるものもある。しかし、降水量や平均気温など取り得る値が実数で無限にある確率変数も存在する。これを連続量と呼ぶが、連続量に対しては確率分布や確率関数をどのように考えればよいだろうか。図 4 の”サイコロルーレット”を例にとって説明しよう。この装

¹ より一般的にはサイコロの目の数を q (一般的なサイコロは $q = 6$)、振る回数を n として、

$$P(x) = \sum_{i=0}^{\lfloor \frac{x-n}{q} \rfloor} \frac{(x-qi-1)!}{(n-1)!(x-qi-n)!} \cdot \frac{n!}{i!(n-i)!} (-1)^i \quad (4)$$

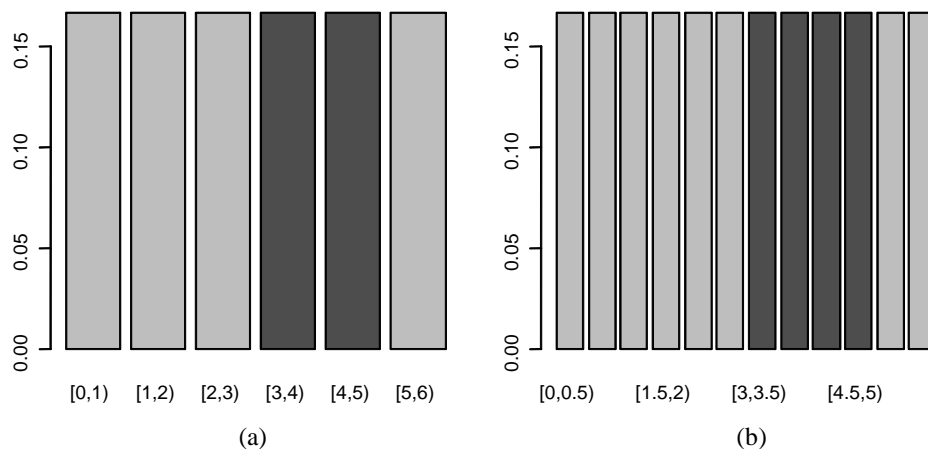


図 5: 連続分布の離散分布への変換

サイコロルーレットから得られる値 X を離散確率分布に近似したもの。 $P(3 \leq X < 5)$ となる面積を濃い色で表してある。(a) 区間の幅を 1 にした場合。(b) 区間の幅をさらに縮めて 0.5 にした場合。

置は円盤と自由に動く針で構成されている。円盤には 0 から 5 までの数が刻んである。針を力強く弾くと、針は何回転もした後、摩擦で停止する。停止したときに針が指し示す実数を X とすれば、針は 0 と 1 の間などどこにでも停止しうするため、 X は連続値となる。確率変数が離散値を取るときと同じような論法で X が例えば 3 を取る確率を論じることができない。針が寸分の狂いもなくちょうど 3 のところで停止することはあり得ないからである (確率は 0)。

そこで連続値を離散値に直すことを考えてみよう。例えば、サイコロルーレットの針が差す数値が 0 以上 1 未満になる確率、1 以上 2 未満になる確率、... というように連続値を各範囲に小分けにして離散値として扱ってはどうか。この場合、各区間の幅 Δx は 1 である。すると図 5(a) に示すように、 X の取り得る範囲は 6 つの領域に分かれ、各範囲内に X が収まる確率は全て $\frac{1}{6}$ となる。またこの確率関数は、p.7 で述べた確率分布の 2 つの性質を満たす。すなわち、

$$P(a \leq X < b) = \sum_{a \leq i < b} P(i \leq X < i+1) \quad (5)$$

なので 2 番目の性質は満たされている。但し、 a, b, i は全て整数で、 $0 \leq a < 6, 0 < b \leq 6$ である。また、

$$P(0 \leq X < 6) = \sum_{0 \leq i < 6} P(i \leq X < i+1) = 1 \quad (6)$$

より、1 の性質も満たす。 X が 3 以上 5 未満になる確率は濃い色の棒の面積の合計で表されることも確認されたい。

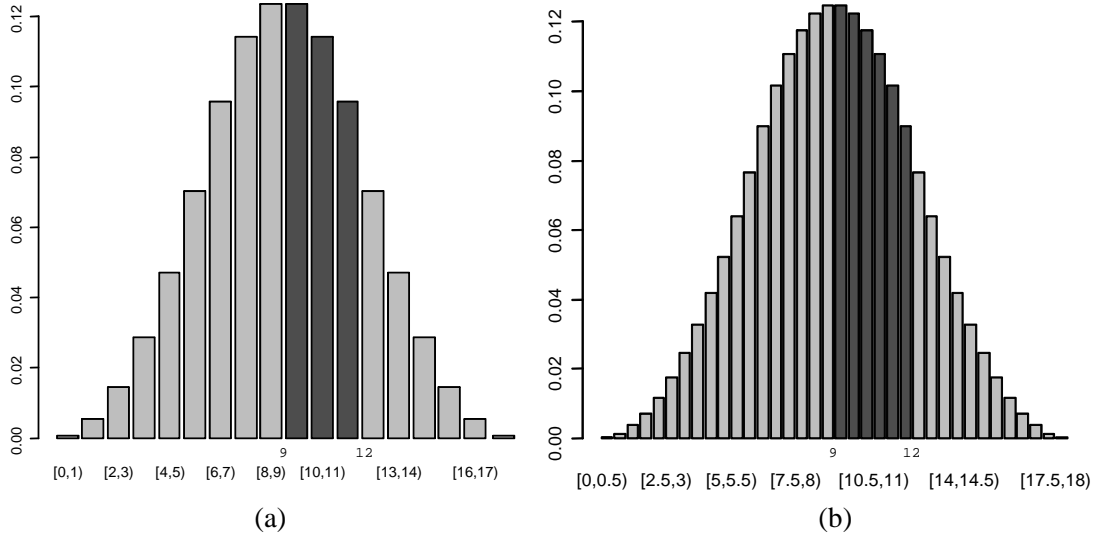


図 6: サイコロルーレットを 3 回まわしたときの合計値の離散確率分布による近似
 合計値が 9 以上 12 未満になる確率に相当する面積を濃い色で表示した。(a) 区間を 1 刻みにして近似した分布。(b) 区間を 0.5 刻みにして近似した分布。

区画をさらに細かくして $\Delta x = 0.5$ 、すなわち 0 以上 0.5 未満、0.5 以上 1 未満、... という範囲に分ければ、図 5(b) のような分布が出来上がる。区間が小さくなれば、 X がその区画に入る確率は当然小さくなってゆく。このとき注意しなければならないのは、 $\Delta x = 0.5$ となっているため、区画 $x \leq X < x + \Delta x$ における棒の高さ=その範囲に X が収まる確率としてしまうと、棒の面積は棒の高さ $\times 0.5$ により、実際の確率の半分になってしまう。そこで棒の面積がその棒の裾の範囲に X が収まる確率にするためには、 $(X$ が入る区画の確率/ Δx) という補正をかける必要がある。これによって

$$P(x \leq X < x + \Delta x) = \text{区画 } x \leq X < x + \Delta x \text{ における棒の高さ} \times \Delta x \quad (7)$$

という重要な関係が保たれる。

これをサイコロルーレットを 3 回回したときの合計値の分布で考えてみよう。区画幅を 1 にしたとき ($\Delta x = 1$) は図 6(a) のような分布になり、区間幅を 0.5 とより細かくしたとき ($\Delta x = 0.5$) は、図 6(b) のようによりスムーズな曲線に近づく。区間幅 Δx を極限まで小さくしていったときに、幅が限りなく 0 に近い棒の高さが一定値に収束するならば、その高さの分布こそが合計値の従う連続確率を表す分布となる。

例えば図 5 の分布は図 7(a) の分布に収束するし、図 6 の分布は、図 7(b) の分布に収束する²。このように収束した分布において、横軸の位置が x のときの棒の高さを表す関数 $f(x)$ のことを確率密度関数と呼ぶ。また確率変数 X が確率密度関数 $f(x)$ に従うことを $X \sim f(x)$ と表す。

²一様分布 $U(0,1)$ から得られる確率変数 X を n 回足した合計値が従う連続分布は $f(x) = \frac{1}{2(n-1)!} \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} (x-k)^{n-1} \text{sgn}(x-k)$ で表される [2, 3]。

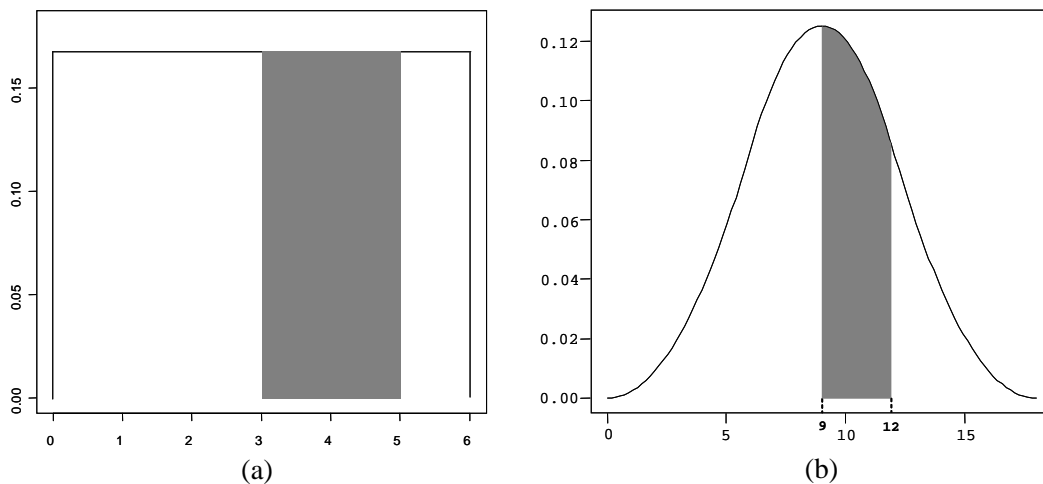


図 7: サイコロルーレットの合計値の連続分布

(a) サイコロルーレットを 1 回回したときに出る値の連続確率分布。値が 3 以上 5 未満になる確率を濃い部分の面積で表した。(b) サイコロルーレットを 3 回回したときに得られる値の合計値の連続確率分布。合計値が 9 以上 12 未満になる確率を濃い面積で表してある。

さてこれらの分布を幅 Δx が限りなく 0 に近い棒の集合と考えるならば、式 7 より、

$$P(x \leq X < x + \Delta x) = f(x)\Delta x \quad (8)$$

となり、また離散確率分布のところで述べた 2 番目の性質から、

$$P(a \leq X < b) = \sum_{a \leq x_i < b} P(x \leq X < x + \Delta x) = \sum_{a \leq x_i < b} f(x_i)\Delta x \quad (9)$$

が成立する。式 9 は $\Delta x \rightarrow 0$ のとき、積分の定義から、

$$P(a \leq X < b) = \int_a^b f(x)dx \quad (10)$$

となる。つまり、確率変数がある範囲の値を取る確率は、確率密度関数の該当面積、すなわち積分で表されるということである。節 1.2 で述べた離散確率分布の 2 つの重要な性質を連続確率分布に当てはめると、確率密度関数を $f(x)$ として、以下ようになる。

1. 確率変数が取りうる全範囲の確率の合計は 1 となる。すなわち、

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad (11)$$

2. X が a 以上、 b 未満となる確率 $P(a \leq X < b)$ は、以下の式で与えられる。

$$P(a \leq X < b) = \int_a^b f(x)dx \quad (12)$$

式 12 より X が x 未満となる確率 $P(X < x)$ は次のように表される。

$$F(x) = P(X < x) = \int_{-\infty}^x f(x)dx \quad (13)$$

この X が x 未満となる確率を表す関数 $F(x)$ を分布関数と呼ぶ。この分布関数を用いて式 12 を書き直すと、以下ようになる。

$$P(a \leq X < b) = F(b) - F(a) \quad (14)$$

$F(x)$ が全ての $x \in R$ において連続であれば³、

$$\frac{dF(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{1}{\Delta x} \int_x^{x+\Delta x} f(x_*)dx_* \quad (15)$$

$\Delta x \rightarrow 0$ の極限において、 $f(x + \Delta x) \approx f(x)$ だから式 15 は、

$$\frac{dF(x)}{dx} = \frac{1}{\Delta x} f(x) \Delta x = f(x) \quad (16)$$

となる。つまり、分布関数を微分したものが確率密度関数となる。

問題 1-2: 確率変数 X の従う確率密度関数 $f(x)$ が

$$f(x) = \begin{cases} x & \text{for } 0 \leq x \leq \sqrt{2} \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

で与えられるとき、 X が 0 以上 1 以下になる確率を求めよ。

1.4 確率密度関数の変数変換

今確率密度関数 $f(x)$ が与えられ、その分布関数は $F(x)$ であるものとし、また x が $x = g(t)$ となる関数 g によって t に対して一意に決まるとする。このとき、 t を変数とした確率密度関数を求めることを考えよう。

$$P(X \leq a) = F(a) = \int_{-\infty}^{x=a} f(x)dx = \lim_{\Delta x \rightarrow 0} \sum_{x_i=-\infty}^{x_i=a} f(x_i)\Delta x = \lim_{\Delta x \rightarrow 0} \sum_{g(t_i)=-\infty}^{g(t_i)=a} f(g(t_i))\Delta x$$

ところで $\Delta t > 0$ に対して $g'(t) > 0$ の場合は $\Delta x = g(t + \Delta t) - g(t)$ 、 $g'(t) < 0$ の場合は、 $\Delta x = -(g(t + \Delta t) - g(t))$ だから、これらをまとめて、 $\Delta x = |g(t + \Delta t) - g(t)|$ と表せる。従って、

³関数 $f(x)$ が $x = a$ において以下の条件を満たすとき、 $f(x)$ が $x = a$ において連続であるという。

1. $f(a)$ が有限確定値。
2. $\lim_{x \rightarrow a} f(x)$ が有限確定値。
3. $f(a) = \lim_{x \rightarrow a} f(x)$ 。

$\epsilon - \delta$ 論法を使って説明すると、どんなに小さい正の数 $\epsilon > 0$ を指定しても、 $|x - a| < \delta$ となる全ての x について常に $|f(x) - f(a)| < \epsilon$ を成立させるような δ が存在する。

$$\begin{aligned}
\lim_{\Delta x \rightarrow 0} \sum_{\substack{g(t_i)=a \\ g(t_i)=-\infty}} f(g(t_i)) \Delta x &= \lim_{\Delta x \rightarrow 0} \sum_{\substack{t_i=g^{-1}(a) \\ t_i=g^{-1}(-\infty)}} f(g(t_i)) |g(t+\Delta t) - g(t)| \\
&= \lim_{\Delta x \rightarrow 0} \sum_{\substack{t_i=g^{-1}(a) \\ t_i=g^{-1}(-\infty)}} f(g(t_i)) \left| \frac{g(t+\Delta t) - g(t)}{\Delta t} \right| \Delta t = \int_{t=g^{-1}(-\infty)}^{t=g^{-1}(a)} f(g(t)) \left| \frac{dg}{dt} \right| dt
\end{aligned}$$

となる。

2 期待値

n 個の値の羅列 x_1, x_2, \dots, x_n の平均 μ は、

$$\mu = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i \quad (18)$$

であり、また分散 σ^2 は、

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (19)$$

と定義される。ちなみに標準偏差 σ は分散 σ^2 の平方根、すなわち、

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (20)$$

で定義される。

次に値が重複している場合を考慮に入れ、 x_1, x_2, \dots, x_i の値を持つ対象がそれぞれ N_1, N_2, \dots, N_n 個観測されたとする。但し、 $\sum_{i=1}^n N_i = N$ とする。これらの対象が持つ値の平均 μ は、 $P_i = \frac{N_i}{N}$ として、

$$\mu = \frac{1}{N} \sum_{i=1}^n x_i N_i = \sum_{i=1}^n x_i \frac{N_i}{N} = \sum_{i=1}^n x_i P_i \quad (21)$$

である。但し、 $\sum_{i=1}^n P_i = 1$ である。一方分散は、

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n (x_i - \mu)^2 N_i = \sum_{i=1}^n (x_i - \mu)^2 \frac{N_i}{N} = \sum_{i=1}^n (x_i - \mu)^2 P_i \quad (22)$$

となる。

これら平均、分散の定義を確率関数にまで拡張することを考えよう。X が取り得る値 x_1, x_2, x_3, \dots とその確率 $P(X = x_1), P(X = x_2), P(X = x_3), \dots$ の積の総和を X の期待値と呼び、 $E_X(X)$ で表す。

$$E_X(X) = \sum_i x_i P(X = x_i) \quad (23)$$

これは式 21 の P_i を $P(X = x_i)$ に置き換えたものである。例えば先に出たサイコロの例では、期待値は 3.5 となる。 E は確率変数の関数 f に対して定義を拡張することが可能であり、 $f(X)$ の X に関する期待値 $E_X(f(X))$ は

$$E_X(f(X)) = \sum_i f(x_i) P(X = x_i) \quad (24)$$

と定義される。なお、どの確率変数に関する期待値であるかがはっきりしている場合は、 $E(X)$ 、 $E(f(X))$ などと省略する。

以上は X が離散値の場合だが、連続値の場合は、 X の取りうる範囲を Δx の大きさに分割すると、

$$E(X) = \sum_i x_i f(x_i) \Delta x \quad (25)$$

となる。 $\Delta x \rightarrow 0$ を取れば式 25 は、

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \quad (26)$$

に収束する。 $E(X)$ を X の平均とも呼ぶ。特に X が密度関数 $f(x)$ に従うとき、 $E(X)$ を $f(x)$ の平均という。

分散 σ^2 は $E((X - E(X))^2)$ で定義され、離散確率の場合は平均を $E(X) = \mu$ としたとき、

$$\sigma^2 = E((X - E(X))^2) = \sum_i (x_i - \mu)^2 P(X = x_i) \quad (27)$$

で表される。

X が $X = X_1 + X_2$ と 2 つの確率変数の和で表されるとき、

$$\begin{aligned} E(X) &= \sum_x x P(X = x) = \sum_x x \sum_{\{(x_1, x_2) | x_1 + x_2 = x\}} P(X_1 = x_1, X_2 = x_2) \\ &= \sum_x \sum_{\{(x_1, x_2) | x_1 + x_2 = x\}} x P(X_1 = x_1, X_2 = x_2) \end{aligned}$$

x の取りうる値について $x_1 + x_2 = x$ となる (x_1, x_2) の和集合は (x_1, x_2) の取りうる値の組み合わせの集合に一致するから

$$\begin{aligned} \sum_x \sum_{\{(x_1, x_2) | x_1 + x_2 = x\}} x P(X_1 = x_1, X_2 = x_2) &= \sum_{x_1} \sum_{x_2} (x_1 + x_2) P(X_1 = x_1, X_2 = x_2) \\ &= \sum_{x_1} \sum_{x_2} x_1 P(X_1 = x_1, X_2 = x_2) + \sum_{x_1} \sum_{x_2} x_2 P(X_1 = x_1, X_2 = x_2) \end{aligned}$$

$\sum_{x_1} P(X = x_1, X = x_2) = P(X = x_2)$, $\sum_{x_2} P(X = x_1, X = x_2) = P(x_1)$ に注意して、

$$\begin{aligned} & \sum_{x_1} \sum_{x_2} x_1 P(X_1 = x_1, X_2 = x_2) + \sum_{x_1} \sum_{x_2} x_2 P(X_1 = x_1, X_2 = x_2) \\ &= \sum_{x_1} x_1 P(X_1 = x_1) + \sum_{x_2} x_2 P(X_2 = x_2) = E_{X_1}(X_1) + E_{X_2}(X_2) \end{aligned}$$

すなわち、 $X = X_1 + X_2$ のとき、

$$E(X) = E(X_1) + E(X_2) \quad (28)$$

となる。この論法を使えば、 $X = X_\alpha + X_\beta$, $X_\alpha = X_1 + X_2$, $X_\beta = X_3 + X_4$ のとき、 $X = X_1 + X_2 + X_3 + X_4$ であり、また $E(X) = E(X_\alpha) + E(X_\beta) = E(X_1) + E(X_2) + E(X_3) + E(X_4)$ となる。確率変数 X_1, X_2, \dots, X_n について、一般に以下の数式が成立する。

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) \quad (29)$$

問題 2-1: 連続確率について、分散を表す式を導け。

問題 2-2: 以下の数式が成立することを示せ。

$$\sigma^2 = E((X - E(X))^2) = E(X^2) - \{E(X)\}^2 = E(X^2) - \mu^2 \quad (30)$$

3 標本分布

ある確率変数 X の性質を調べたいときに、 X がどのような値を取るかを次々に見ていく方法がある。例えば X を日本人 20 歳男子の身長だとしよう。確実に分布を得るためには、日本人 20 歳男子全ての身長を測る方法があるだろう。しかしながらこれは時間・費用・労力の面から効率の良い方法とは言えないし、多くの場合現実的ではないだろう。そこで多くの場合、無作為に標本を取り、その標本の性質から全体の性質を推測することが行われる。標本のことから母数のことがどれほどのことがどれほどの精度で分かるだろうか？

まずは母数の身長の平均値 μ とその標本 i の標本中の平均 \bar{x}_i との関係を調べよう。例えば n 人分の標本を採れば、その標本に対して標本平均 \bar{x} と標本分散 s^2 を計算できる。この \bar{x} や s^2 はどのような性質を持っているだろうか。まずは \bar{x} の方から見てみよう。 \bar{x} の性質を調べるためには、 \bar{x} 自体の分布、すなわち図 8 に示すように、 n 人の標本 (この例では $n = 7$) を採るという操作を繰り返して行ったとき、 \bar{x} がどのような分布に従うかを調べるのがよい。ある標本について n 人分の身長を X_1, X_2, \dots, X_n (但し X_i は確率変数とみなす) とすれば、その標本平均 \bar{X} は、

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (31)$$

である。ここで \bar{X} もまた確率変数である。標本平均 \bar{X} の平均 $E(\bar{X})$ は $E(X_i) = \mu$ の関係に注意しつつ、式 29 の関係を使うと、

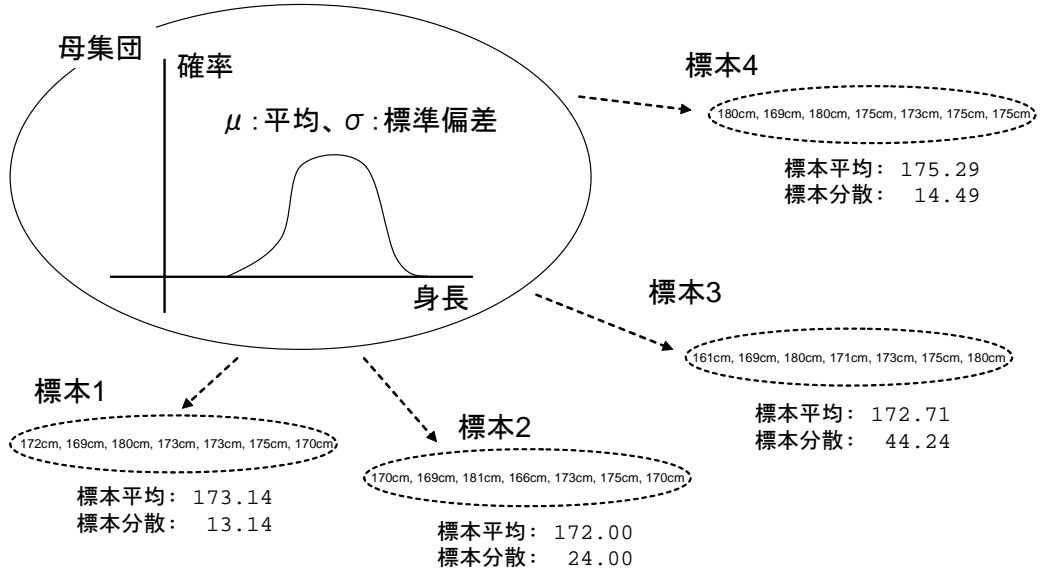


図 8: 身長の平均の例

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n} (E(X_1) + E(X_2) + \cdots + E(X_n)) = \frac{1}{n}(n\mu) = \mu \quad (32)$$

となり、標本平均 \bar{X} の分布の平均 $E(\bar{X})$ は母集団の分布の平均 $E(X)$ に一致することが示された。
標本平均の分散 $\sigma_{\bar{x}}^2$ は、

$$\begin{aligned} \sigma_{\bar{x}}^2 &= E((\bar{X} - \mu)^2) = E\left\{\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - \mu\right)^2\right\} \\ &= \frac{1}{n^2} E\left\{((X_1 - \mu) + (X_2 - \mu) + \cdots + (X_n - \mu))^2\right\} \\ &= \frac{1}{n^2} \left\{ \sum_i E\{(X_i - \mu)^2\} \right\} + \frac{1}{n^2} \left\{ \sum_{i \neq j} E\{(X_i - \mu)(X_j - \mu)\} \right\} \end{aligned} \quad (33)$$

式 33 の最後の等号より後ろの部分の第 2 項は、

$$\begin{aligned} E\{(X_i - \mu)(X_j - \mu)\} &= \sum_{x_i} \sum_{x_j} (x_i - \mu)(x_j - \mu) P(X = x_i) P(X = x_j) \\ &= \sum_{x_i} (x_i - \mu) P(X = x_i) \sum_{x_j} (x_j - \mu) P(X = x_j) = E(X - \mu) \cdot E(X - \mu) = 0 \end{aligned} \quad (34)$$

また $E(X_i - \mu)^2 = \sigma^2$ より、式 33 は

$$\sigma_{\bar{x}}^2 = E((\bar{X} - \mu)^2) = \frac{1}{n^2} \left\{ \sum_i E\{(X_i - \mu)^2\} \right\} = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \quad (35)$$

となる。サンプル数 n が多くなるに従って、標本平均の分散は小さくなっていくことが分かる。標本平均の標準偏差 $\sigma_{\bar{x}}$ は標本平均の標準誤差と呼ばれ、

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (36)$$

で表される。

問題 3-1: 式 35 の導出過程を参考に、 $E(\bar{X}^2)$ を n, μ, σ を用いて表せ。

問題 3-2: 問題 3-1 の結果および式 30 を活用して、 $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ の平均 $E\{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\}$ を n, μ, σ を用いて表せ。

問題 3-3: 問題 3-1 の結果および式 30 を活用して、標本分散 s^2 の平均 $E(s^2) = E\{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\}$ を n, μ, σ を用いて表せ。

4 二項分布

サイコロの目について他の確率現象を考えてみよう。サイコロを 3 回振ったときに 5 の目が 2 回出る確率はどれくらいだろうか。サイコロを 1 回目に振ったとき、2 回目に振ったとき、そして 3 回目に振ったときの目をそれぞれ X_1, X_2, X_3 で表すと、サイコロを振る試行は毎回独立と考えられるので (つまり 1 回目の結果が 2 回目に影響することはない)、

$$P(X_1 \wedge X_2 \wedge X_3) = P(X_1)P(X_2)P(X_3) \quad (37)$$

となる。これを元に以下のステップで 3 回中 2 回 5 が出る確率を考えることができる。

1. 5 の目が出る確率は、 $P(X_i = 5) = \frac{1}{6}$ ($1 \leq i \leq 3$) である。
2. 5 以外の目が出る確率は、 $P(X_i \neq 5) = \frac{5}{6}$ ($1 \leq i \leq 3$) である。
3. サイコロの目が 2 回 5 になる組み合わせは、 $X_1 = 5, X_2 = 5, X_3 \neq 5$ または $X_1 = 5, X_2 \neq 5, X_3 = 5$ または $X_1 \neq 5, X_2 = 5, X_3 = 5$ の 3 通りである。
4. いずれの場合も 5 が 2 回、5 以外が 1 回なので、各確率は式 37 より $(\frac{1}{6})^2 \frac{5}{6}$ である。
5. さらに各事象は排反なので、結局 5 の目が 2 回出る確率は $3 \times (\frac{1}{6})^2 \frac{5}{6}$ となる。

これを一般に拡張して考えよう。今事象 a が起こる確率が p の試行を n 回繰り返すことを考えよう。 i 回目 $1 \leq i \leq n$ の試行結果を X_i とすると、各試行が独立であるとすれば、

$$P(X_1 \wedge P_2 \wedge \cdots \wedge P_n) = P(X_1)P(X_2) \cdots P(X_n) \quad (38)$$

である。

1. 各試行において、事象 a が起こる確率は $P(X = a) = p$ である。
2. 各試行において、事象 a が起こらない確率は $P(X \neq a) = 1 - p$ である。
3. n 回の試行の中で事象 a が x 回起きるときの $X_{a:1} = a, X_{a:2} = a, \dots, X_{a:x} = a$ の組合せの個数を ${}_nC_x$ と表す⁴。
4. いずれの場合も a が x 回、 a 以外が $n - x$ 回なので、各確率は $p^x(1 - p)^{n-x}$ である。
5. さらに各事象は排反なので、結局事象 a が x 回起こる確率は $p^x(1 - p)^{n-x} {}_nC_x$ である。

次に n 個の対象のなかから x 個の対象を選ぶ組み合わせの ${}_nC_x$ がいくつになるかを考えよう。最初に n 個の中からの最初の 1 個の選び方はもちろん n 通りである。最初の選択肢のそれぞれについて、残りの $n - 1$ 個からの次の 2 個目の選び方は $n - 1$ 通りであるため、1 個目と 2 個目の組み合わせの数は $n(n - 1)$ 通りとなる。さらに 1 個目と 2 個目の選択肢のそれぞれについて、残りの中から $n - 2$ 通りの選び方があるため、1 個目、2 個目、3 個目の選び方は $n(n - 1)(n - 2)$ 通りとなる。この考え方を進めると、1, 2, 3, \dots , x 個目の選び方は全部で $n(n - 1)(n - 2) \cdots (n - x + 1)$ 通りとなる。

今 n 個の対象に 1, 2, \dots , n と番号を振る。すると例えば $n = 4$ で $x = 3$ のとき、選び方は、(1, 2, 3), (1, 2, 4), (1, 3, 2), (1, 3, 4), (1, 4, 2), (1, 4, 3), (2, 1, 3), (2, 1, 4), (2, 3, 1), (2, 3, 4), (2, 4, 1), (2, 4, 3), (3, 1, 2), (3, 1, 4), (3, 2, 1), (3, 2, 4), (3, 4, 1), (3, 4, 2), (4, 1, 2), (4, 1, 3), (4, 2, 1), (4, 2, 3), (4, 3, 1), (4, 3, 2) の $4 \times 3 \times 2 = 24$ 通りとなる。ところがよく見てみると、実は 1, 2, 3 と 1, 3, 2 が別の選び方として数えられている。1, 2, 3 を含む組み合わせは $3 \times 2 \times 1 = 6$ 通りある。これは全ての (互いに異なる) 組み合わせについて成立するため、組み合わせで考えた場合、結局 $\frac{4 \times 3 \times 2}{3 \times 2 \times 1} = 4$ 通りであることが分かる。これを一般化すると、1, 2, 3, \dots , x には全部で $x(x - 1) \cdots 2 \cdot 1 = x!$ の並び方があるため、以下が成立する。

$${}_nC_x = \frac{n(n - 1)(n - 2) \cdots (n - x + 1)}{x!} = \frac{n!}{x!(n - x)!} \quad (39)$$

以上により、事象 a が起こる確率が p の試行を n 回繰り返したとき、 a が x 回観測されときの確率分布 $B(n, p)$ の関数 $P(x)$ は、

$$P(x) = p^x(1 - p)^{n-x} {}_nC_x = p^x(1 - p)^{n-x} \frac{n!}{x!(n - x)!} \quad (40)$$

となる。

図 9 にこの分布の例を 3 つ示す。 n が大きくなるにつれ、分布が全体的になめらかかつ左右対称に近付いてゆくことが分かる。

平均 μ は以下のように計算される。

$$\mu = \sigma \sum_{x=0}^n x {}_nC_x p^x (1 - p)^{n-x} = \sum_{x=1}^n x \cdot \frac{n!}{x!(n - x)!} p^x (1 - p)^{n-x}$$

⁴ ${}_nC_x$ を $\binom{n}{x}$ と記述している書籍も多数ある。

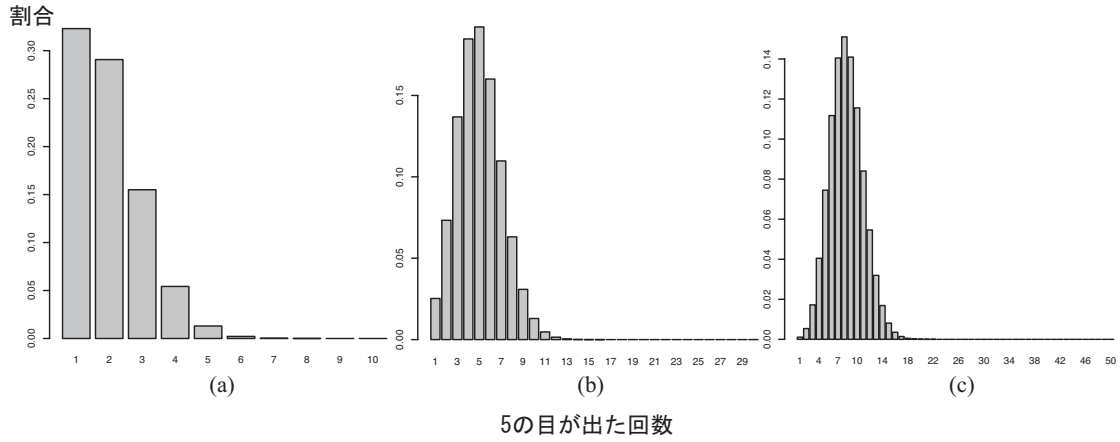


図 9: サイコロを n 回振ったときに 5 の目が出る回数が従う分布

(a) $n = 10$ 、 (b) $n = 30$ 、 (c) $n = 50$ のときの二項分布 $B(n, p)$ を示した。

$$\begin{aligned}
 &= np \sum_{x=1}^{n-1} \frac{(n-1)!}{(x-1)!(n-1-(x-1))!} p^{x-1} (1-p)^{n-1-(x-1)} \\
 &= np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y (1-p)^{n-1-y} = np \sum_{y=0}^{n-1} {}_{n-1}C_y p^y (1-p)^{n-1-y} \\
 &= np(p+1-p)^{n-1} = np
 \end{aligned} \tag{41}$$

また分散 σ^2 は以下のように計算される。

$$\begin{aligned}
 \sigma^2 &= E(X^2) - \{E(X)\}^2 = E\{X(X-1) + X\} - \mu^2 \\
 &= E\{X(X-1)\} + \mu - \mu^2
 \end{aligned} \tag{42}$$

ここで、

$$\begin{aligned}
 E\{X(X-1)\} &= \sum_{x=0}^n x(x-1) {}_n C_x p^x (1-p)^{n-x} \\
 &= \sum_{x=2}^n x(x-1) \cdot \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
 &= n(n-1)p^2 \sum_{x=2}^n \frac{(n-2)!}{(x-2)!(n-2-(x-2))!} p^{x-2} (1-p)^{n-2-(x-2)} \\
 &= n(n-1)p^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{y!(n-2-y)!} p^y (1-p)^{n-2-y} \\
 &= n(n-1)p^2 \sum_{y=0}^{n-2} {}_{n-2}C_y p^y (1-p)^{n-2-y} = n(n-1)p^2(p+(1-p))^{n-2} \\
 &= n(n-1)p^2
 \end{aligned}$$

であるから、これを式 42 に代入して、

$$\sigma^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p) \quad (43)$$

となる。

5 正規分布

節 4 で説明した二項分布において、 n が大きくなると、 $P(x)$ はどのような分布になっていくのだろうか。これを以下で説明しよう。まず x を以下の式で z に正規化する。

$$z = \frac{(x - np)}{\sqrt{np(1-p)}} \quad (44)$$

このとき z が従う分布を $g(z)$ とし、 $g(z)$ を求めよう。 x は $0, 1, 2, \dots$ と離散的な値をとるので、 x の変化量 Δx の最小値は 1 である。このとき、 z の変化量 Δz は常に

$$\Delta z = \frac{1}{\sqrt{np(1-p)}} \quad (45)$$

となる。式 44 を式 45 で割って変形すると、

$$x = np + \frac{z}{\Delta z} \quad (46)$$

が得られる。式 40 の x に $x+1$ を代入すると、

$$P(x+1) = {}_nC_{x+1} \cdot p^{x+1}(1-p)^{n-x-1} \quad (47)$$

ここで、

$${}_nC_x = \frac{n!}{x!(n-x)!}$$

を使って式 47 を式 40 で割り、式 48 を代入すると、

$$\frac{P(x+1)}{P(x)} = \frac{(n-x)p}{(x+1)(1-p)} \quad (48)$$

となる。さてここで $P(x)$ と $g(z)$ の関係を考えよう。 z と x は式 44 のような関係で結ばれている。さて一般的に離散的な値を取る確率変数 X が確率密度関数 P_d に従うとき、 $X = x$ である確率は $P_d(x)\Delta x$ で求められる。但し、 Δx は確率変数 X が取り得る値の中で $> x$ を満たす最小のものと x との差である⁵。従って $g(z)$ が確率密度関数であるためには、 $g(z)\Delta z$ が確率 $P(x)\Delta x = P(x)$ を表さなければならない。すなわち、

$$g(z) = \frac{1}{\Delta z} P(x) \quad (49)$$

⁵ $x \leq X < x + \Delta x$ において、 $P_d(X)$ は一定と考える。

および

$$g(z + \Delta z) = \frac{1}{\Delta z} P(x + 1) \quad (50)$$

が成立する。式 50 を式 49 で割り、式 48 を代入すると、

$$\frac{g(z + \Delta z)}{g(z)} = \frac{P(x + 1)}{P(x)} = \frac{(n - x)p}{(x + 1)(1 - p)} \quad (51)$$

式 51 に 46 を代入して x を消去すると、

$$\frac{g(z + \Delta z)}{g(z)} = \frac{(n - np - z/\Delta z)p}{(np + z/\Delta z + 1)(1 - p)} = \frac{np(1 - p) - pz/\Delta z}{np(1 - p) + (1 - p)(z/\Delta z + 1)} \quad (52)$$

式 45 より $np(1 - p) = \frac{1}{\Delta z^2}$ なので、これを上の式に代入して、

$$\frac{g(z + \Delta z)}{g(z)} = \frac{\frac{1}{\Delta z^2} - p\frac{z}{\Delta z}}{\frac{1}{\Delta z^2} + (1 - p)(\frac{z}{\Delta z} + 1)} = \frac{1 - pz\Delta z}{1 + (1 - p)\{z\Delta z + \Delta z^2\}} \quad (53)$$

となる。式 53 を使うと、

$$\begin{aligned} \frac{g(z + \Delta z) - g(z)}{\Delta z} &= \left(\frac{g(z + \Delta z)}{g(z)} - 1 \right) \frac{g(z)}{\Delta z} \\ &= \frac{1 - pz\Delta z - 1 - (1 - p)(z\Delta z + \Delta z^2)}{1 + (1 - p)(z\Delta z + \Delta z^2)} \frac{g(z)}{\Delta z} \\ &= \frac{-z\Delta z - (1 - p)\Delta z^2}{1 + (1 - p)(z\Delta z + \Delta z^2)} \frac{g(z)}{\Delta z} = \frac{-z - (1 - p)\Delta z}{1 + (1 - p)(z\Delta z + \Delta z^2)} g(z) \end{aligned} \quad (54)$$

が得られる。式 45 より $n \rightarrow \infty$ のとき、 $\Delta z \rightarrow 0$ となる。このとき、式 54 の極限は

$$\frac{dg(z)}{dz} = -zg(z) \quad (55)$$

となる。これを解くと、

$$g(z) = Ce^{-\frac{1}{2}z^2}$$

が得られる。但し C は定数である。以下、この C を定めよう。 $g(z)$ は確率密度関数でありまた偶関数⁶なので、

$$\int_{-\infty}^{+\infty} g(z) dz = \int_{-\infty}^{+\infty} Ce^{-\frac{1}{2}z^2} dz = 2\sqrt{2}C \int_0^{\infty} e^{-z^2} dz = 1 \quad (56)$$

が成立する。式 56 の z を y に変えても式全体の値は変わらないから、

$$\left(\int_0^{+\infty} e^{-z^2} dz \right)^2 = \left(\int_0^{+\infty} e^{-z^2} dz \right) \left(\int_0^{+\infty} e^{-y^2} dy \right) = \int_0^{+\infty} \int_0^{+\infty} e^{-(z^2+y^2)} dy dz \quad (57)$$

⁶ $f(x) = f(-x)$ が常に成立する関数を偶関数と呼ぶ。これに対し、 $f(x) = -f(-x)$ が常に成立する関数を奇関数と呼ぶ。

ここで $y = zu$ とおくと⁷、 $\frac{dy}{du} = z$ すなわち $dy = zdu$ より、

$$\begin{aligned} \int_0^{+\infty} \int_0^{+\infty} e^{-(z^2+y^2)} dydz &= \int_0^{+\infty} \left(\int_0^{+\infty} e^{-(1+u^2)z^2} zdu \right) dz \\ &= \int_0^{+\infty} \int_0^{+\infty} e^{-(1+u^2)z^2} z dz du = \int_0^{+\infty} \left[-\frac{1}{2(1+u^2)} \cdot e^{-(1+u^2)z^2} \right]_0^{+\infty} du \\ &= \int_0^{+\infty} \frac{du}{2(1+u^2)} = \frac{\pi}{4} \end{aligned} \quad (58)$$

となり、 $C = \frac{1}{\sqrt{2\pi}}$ が導かれる。結局 z の確率密度関数は、

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

になることが導かれた。この確率密度関数に従う分布を標準正規分布と呼ぶ。

$N(\mu, \sigma^2)$ に従う正規分布 $f(x)$ は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (59)$$

で表される。

問題 5-1: 三角関数の関係式 $\frac{\sin \theta}{\cos \theta} = \tan \theta$, $\sin^2 \theta + \cos^2 \theta = 1$, $\frac{d \tan \theta}{d \theta} = \frac{1}{\cos^2 \theta}$ を駆使して、式 58 の最後の部分 $\int_0^{+\infty} \frac{du}{2(1+u^2)} = \frac{\pi}{4}$ が成立することを示せ。

6 ポワソン分布

二項分布から導かれる分布は他にもある。身近な例でこれを説明しよう。今ある地方で深夜の 1 時間の間に流れ星が平均 3 個観測されるものとする。各流れ星は一瞬しか観測されないものとし、1 時間のうちのどの時間帯に観測されるかは全て同様に確からしいとする。また、流れ星の出現という事象は複数の流れ星に関して互いに独立であるとする。

ここで図 10(a) に示すように、60 分を 5 個 ($n = 5$) の時間帯、すなわち 12 分ずつに分けよう。ここで仮定をもうひとつ置く。各時間帯で流れ星が 2 つ以上観測されることはない、すなわち各時間帯で流れ星は全く観測されないか、1 個だけ観測されるかのどちらかである。各時間帯で流れ星が観測される確率を $p(n)$ とすれば、60 分の間に流れ星が観測される個数は二項分布 $B(5, p(5))$ に従う。ここで平均を λ とおけば、 $\lambda = 5 \cdot p(5) = 3$ である。

分割数 n を増やし、図 10(b,c) に示すように、 $n = 10, 20$ と増やしていったらどうだろうか。この場合も $\lambda = n \cdot p(n) = 3$ で一定となる。さらに n を増やしたとき、この二項分布はどのような分布に近づいていくかを計算すると以下ようになる。60 分の間に流れ星が観測される数を x 、その確率を $P(x)$ とおくと、 $n \rightarrow \infty$ のとき、

⁷ $y = zu$ の制約の中でも、 u の値を調整すれば、 z, y は正のあらゆる組み合わせの値を取り得ることに注意。式の中に 3 つの変数全て (z, y, u) が登場すると、これらの 3 つの変数の取り得る値の組み合わせは限定される。

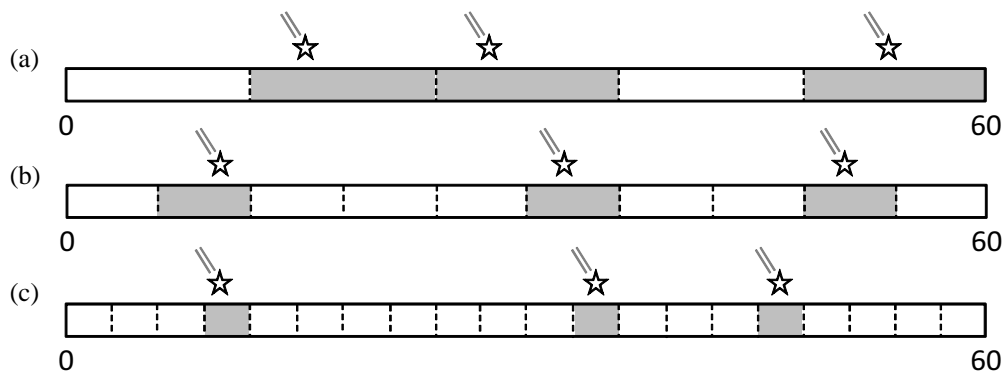


図 10: 60 分の間に流れ星が観測される回数の統計的考察

60 分を (a)5 つの時間帯に分けた場合、(b)10 個の時間帯分けた場合、(c)20 個の時間帯に分けた場合を示した。

$$\begin{aligned}
 P(x) &= \lim_{n \rightarrow \infty} {}_n C_x p(n)^x (1 - p(n))^{n-x} \\
 &= \lim_{n \rightarrow \infty} \frac{n!}{x!(n-x)!} \cdot \left(\frac{\lambda}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{x!} \cdot \frac{n(n-1) \cdots (n-x+1)}{nn \cdots n} \cdot \lambda^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
 &= \lim_{n \rightarrow \infty} \frac{\lambda^x}{x!} \cdot \underbrace{1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right)}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{e \rightarrow -\lambda} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_{\rightarrow 1} \\
 &= \frac{\lambda^x}{x!} e^{-\lambda}
 \end{aligned} \tag{60}$$

となる。このように確率関数 $P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$ に従う確率分布をポワソン分布 (Poisson distribution) という。

7 テイラー展開

7.1 平均値の定理

ロル (Rolle) の定理: $f(x)$ が区間 $[a, b]$ で連続、 (a, b) で微分可能であり、 $f(a) = f(b) = 0$ なら、 $f'(c) = 0$ ($a < c < b$) となる c が少なくとも 1 つ存在する。

証明: $f(x)$ が定数なら、 $f'(x) = 0$ となり、定理は明らかである。それ以外に関しては、閉区間 $[a, b]$ で $f(x)$ には最大値 M と最小値 m が存在する。 $M \neq f(a)$ なら、最大値を与える点を $x = c$ とすると、 $M = f(c)$ 。十分小さい h に対し $f(c+h) - f(c) \leq 0$ だから、

$$\begin{aligned}\frac{f(c+h)-f(c)}{h} &\leq 0 & (h > 0) \\ \frac{f(c+h)-f(c)}{h} &\geq 0 & (h < 0)\end{aligned}$$

$f(x)$ は微分可能だから、 $h \rightarrow 0$ の極限を取れば両者は $f'(c) = 0$ を与えるが、 $f'(c) \leq 0$ かつ $f'(c) \geq 0$ でなければならないから、 $f'(c) = 0$ 。また $M = f(a)$ のとき、 (a, b) に最小値 m を与える $x = c'$ が存在するから、上と同様にして、 $f'(c) \geq 0, f'(c) \leq 0$ でなければならない、 $f'(c') = 0$ となる c' が (a, b) 内に存在する。

平均値の定理: $f(x)$ が区間 $[a, b]$ で連続で区間 (a, b) で微分可能ならば、

$$f(b) - f(a) = f'(c)(b - a) \quad (61)$$

となる c ($a < c < b$) が少なくとも 1 つ存在する。

証明:

$$F(x) = f(b) - f(x) - \frac{f(b) - f(a)}{b - a}(b - x)$$

を考えれば、 $F(x)$ は $[a, b]$ で連続、 (a, b) で微分可能である。 $F(a) = F(b)$ だから、 $F(x)$ のロルの定義が適用できて、 $F'(c) = 0$ となる c が少なくとも区間 (a, b) に 1 つは存在する。

$$F'(x) = -f'(x) + \frac{f(b) - f(a)}{b - a}$$

より、 $F'(c) = 0$ とおいて、

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

(a, b) 内の点 c は $c = a + \theta(b - a)$ ($0 < \theta < 1$) と表すこともでき、式 61 は、

$$f(b) - f(a) = f'(a + \theta(b - a))(b - a) \quad (62)$$

と書くことができる。

7.2 テイラーの定理

定理: $f(x), f'(x), \dots, f^{(n-1)}(x)$ は $[a, b]$ で連続で (a, b) で $f^{(n)}(x)$ が存在すれば、

$$\begin{aligned}
f(b) &= f(a) + \frac{f'(a)}{1!}(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \\
&\quad \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b-a)^n \\
&\quad (a < c < b)
\end{aligned} \tag{63}$$

となるような c が存在する。

証明：

$$\begin{aligned}
f(b) &= f(a) + \frac{f'(a)}{1!}(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \\
&\quad \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{k}{n!}(b-a)^n
\end{aligned} \tag{64}$$

とおく。 k 以外の変数が固定されていても、 k の値を調整すれば、式 64 を成立させることができる。
右辺の a を x に置き換え、 $f(b)$ から引くと、

$$\begin{aligned}
F(x) &= f(b) - f(x) - \frac{f'(x)}{1!}(b-x) - \frac{f''(x)}{2!}(b-x)^2 - \\
&\quad \cdots - \frac{f^{(n-1)}(x)}{(n-1)!}(b-x)^{n-1} + \frac{k}{n!}(b-x)^n
\end{aligned} \tag{65}$$

$F(a) = 0, F(b) = 0$ であり、 Rolle の定理から $F'(c), a < c < b$ となる c が存在する。

$$F'(c) = -\frac{f^{(n)}(c)}{(n-1)!}(b-c)^{n-1} + \frac{k}{(n-1)!}(b-c)^{n-1} = 0 \tag{66}$$

より $k = f^{(n)}(c)$ を得る。

区間 $[a, b]$ は任意だから、 $a = 0, b = x$ とおけば次のマクローリン (Maclaurin) 展開が得られる。

$$\begin{aligned}
f(x) &= f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots \\
&\quad + \frac{f^{(n-1)}(0)}{(n-1)!}x^{n-1} + \frac{f^{(n)}(\theta x)}{n!}x^n \quad (0 < \theta < 1)
\end{aligned} \tag{67}$$

8 偏微分

n 個の独立変数を持つ関数 $f(x, x_2, x_3, \dots, x_n)$ に対して定義される

$$f_x(x, x_2, \dots, x_n) = \frac{\partial f(x, x_2, \dots, x_n)}{\partial x} = \frac{\partial f}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, x_2, \dots, x_n) - f(x, x_2, \dots, x_n)}{\Delta x} \tag{68}$$

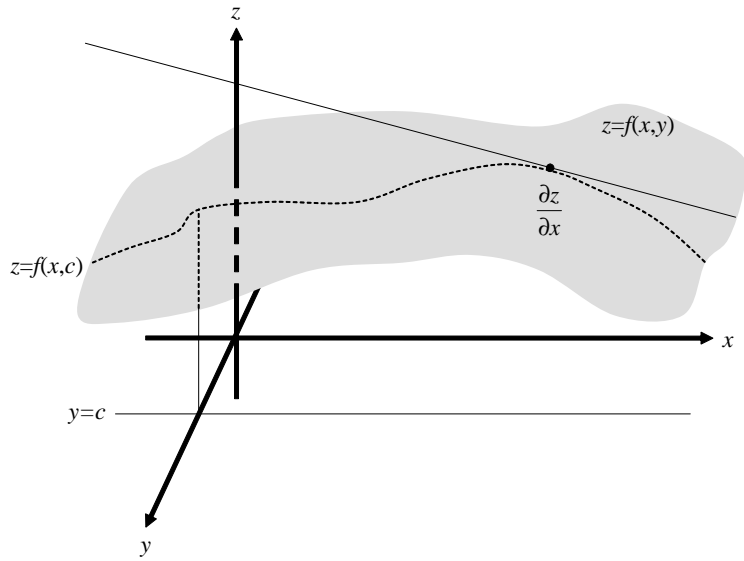


図 11: $z = f(x, y)$ の偏微分の幾何学的意味

を f の x による偏微分と呼ぶ。

今簡単のために、2 変数の関数 $f(x, y)$ を考える。その幾何学的な意味を図 11 に示す。

$$\begin{aligned}\Delta f &= f(x + \Delta x, y + \Delta y) - f(x, y) \\ &= f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) + f(x, y + \Delta y) - f(x, y)\end{aligned}\tag{69}$$

ここで平均値の定理を用いれば、

$$\begin{aligned}f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y) &= \Delta x \cdot f_x(x + \theta_1 \Delta x, y + \Delta y) \\ f(x, y + \Delta y) - f(x, y) &= \Delta y \cdot f_y(x, y + \theta_2 \Delta y) \\ (0 < \theta_1, \theta_2 < 1)\end{aligned}$$

と書けるから、式 69 は

$$\Delta f = \Delta x \cdot f_x(x + \theta_1 \Delta x, y + \Delta y) + \Delta y \cdot f_y(x, y + \theta_2 \Delta y)\tag{70}$$

ここで、

$$\begin{aligned}\lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} \frac{\partial f(x + \theta_1 \Delta x, y + \Delta y)}{\partial x} &= \frac{\partial f(x, y)}{\partial x} \\ \lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} \frac{\partial f(x, y + \theta_2 \Delta y)}{\partial y} &= \frac{\partial f(x, y)}{\partial y}\end{aligned}$$

を使って、

$$\begin{aligned}\frac{\partial f(x + \theta_1 \Delta x, y + \Delta y)}{\partial x} &= \frac{\partial f(x, y)}{\partial x} + \epsilon_1 \\ \frac{\partial f(x, y + \theta_2 \Delta y)}{\partial y} &= \frac{\partial f(x, y)}{\partial y} + \epsilon_2\end{aligned}\tag{71}$$

とおけば、 $\Delta x \rightarrow 0, \Delta y \rightarrow 0$ のとき、 $\epsilon_1 \rightarrow 0, \epsilon_2 \rightarrow 0$ 。従って、

$$\Delta f \approx \frac{\partial f(x, y)}{\partial x} \Delta x + \frac{\partial f(x, y)}{\partial y} \Delta y + \epsilon_1 \Delta x + \epsilon_2 \Delta y\tag{72}$$

$\Delta x \rightarrow 0, \Delta y \rightarrow 0$ のとき、

$$df = \frac{\partial f(x, y)}{\partial x} dx + \frac{\partial f(x, y)}{\partial y} dy\tag{73}$$

これを f の全微分という。

$f(x, y)$ が (a, b) の近傍で定義されており、 (a, b) の近くの任意の (x, y) に対して常に $f(x, y) < f(a, b)$ が成立するとき、 $f(x, y)$ は点 (a, b) で極大をとるといい、 $f(a, b)$ を極大値という。また逆に $f(x, y) > f(a, b)$ であれば、 $f(x, y)$ は $f(a, b)$ で極小をとるといい、 $f(a, b)$ を極小値という。極大値、極小値を総称して極値という。

定理: $f(x, y), f_x(x, y), f_y(x, y)$ は (a, b) 近くで連続で $f(x, y)$ が点 (a, b) で極値をとるならば、 $f_x(a, b) = 0, f_y(a, b) = 0$ である。

証明: $F(x) = f(x, b)$ とおけば、 $F(x)$ は $x = a$ で極値を持つから、 $F'(a) = 0$ 、つまり $f_x(a, b) = 0$ 。また $G(y) = f(a, y)$ は $y = b$ で極値を持つから、 $G'(b) = 0$ 。つまり、 $f_y(a, b) = 0$ である。

9 行列式

9.1 置換

有限集合 ω から ω 自身への 1 対 1 の写像を ω の置換という。特に $M = \{1, 2, \dots, n\}$ とするとき、 M の置換を n 文字の置換という。置換 $\sigma : M \rightarrow M$ を

$$\sigma = \left(\begin{array}{cccc} 1 & 2 & \cdots & n \\ \Downarrow & \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{array} \right)\tag{74}$$

と表す。

$$I = \left(\begin{array}{cccc} 1 & 2 & \cdots & n \\ \Downarrow & 1 & 2 & \cdots & n \end{array} \right)\tag{75}$$

を恒等置換という。また 2 つの置換

$$\sigma = \left(\begin{array}{cccc} \Downarrow & 1 & 2 & \cdots & n \\ & \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{array} \right), \quad \tau = \left(\begin{array}{cccc} \Downarrow & 1 & 2 & \cdots & n \\ & \tau(1) & \tau(2) & \cdots & \tau(n) \end{array} \right) \quad (76)$$

に対し、 σ と τ の合成写像

$$\begin{aligned} \tau\sigma &= \left(\begin{array}{cccc} \Downarrow & \sigma(1) & \sigma(2) & \cdots & \sigma(n) \\ & \tau(\sigma(1)) & \tau(\sigma(2)) & \cdots & \tau(\sigma(n)) \end{array} \right) \left(\begin{array}{cccc} \Downarrow & 1 & 2 & \cdots & n \\ & \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{array} \right) \\ &= \left(\begin{array}{cccc} \Downarrow & 1 & 2 & \cdots & n \\ & \tau(\sigma(1)) & \tau(\sigma(2)) & \cdots & \tau(\sigma(n)) \end{array} \right) \end{aligned} \quad (77)$$

を σ と τ の積と呼ぶ。また置換 $\sigma = \left(\begin{array}{cccc} \Downarrow & 1 & 2 & \cdots & n \\ & \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{array} \right)$ に対して、

$$\sigma^{-1} = \left(\begin{array}{cccc} \Downarrow & \sigma(1) & \sigma(2) & \cdots & \sigma(n) \\ & 1 & 2 & \cdots & n \end{array} \right) = \left(\begin{array}{cccc} \Downarrow & 1 & 2 & \cdots & n \\ & \sigma^{-1}(1) & \sigma^{-1}(2) & \cdots & \sigma^{-1}(n) \end{array} \right) \quad (78)$$

を σ の逆置換という。さらに i, j を異なる 2 つの文字とすると、置換 σ によって

$$\sigma(i) = j, \sigma(j) = i, \quad \text{その他の } M \text{ の文字は動かない} \quad (79)$$

のとき、すなわち、 i と j だけが入れ替わるとき、 σ は互換であるといい、

$$\sigma = (i, j) \quad (80)$$

で表す。そして、 i_1, \dots, i_r を M の互いに異なる文字とすると、置換 σ によって

$$\sigma(i_1) = i_2, \sigma(i_2) = i_3, \dots, \sigma(i_r) = i_1, \quad \text{その他の文字は動かない} \quad (81)$$

とすると、 σ を r 次の巡回置換といい、

$$\sigma = (i_1, i_2, \dots, i_r) \quad (82)$$

で表す。

定理: 任意の置換は互換の積で表せる。

証明: 置換 σ が恒等置換 I ならば、それは 0 個の互換の積と考えられるので、以下、 $\sigma \neq I$ について考える。まず σ が巡回置換の積で表されることを示す。 $\sigma(i_1) \neq i_1$ となる i を 1 つとり、 $i, \sigma(i_1), \sigma^2(i_1), \dots$ の中で最初に $\sigma^{r_1}(i_1) = i_1$ となる自然数 r_1 をとる。このとき、以下の巡回置換 σ_1 を考える。

$$\sigma_1 = (i_1, \sigma(i_1), \dots, \sigma^{r_1-1}(i_1)) \quad (83)$$

このとき $\sigma = \sigma_1$ ならば、 σ が巡回置換 1 つの積と考えられる。 $\sigma \neq \sigma_1$ ならば、 $i_1, \sigma(i_1), \dots, \sigma_{r_1-1}(i_1)$ のどれとも異なる i_2 で、 $\sigma(i_2) \neq i_2$ を満たす文字 i_2 がある。上と同様にして、以下の巡回置換が得られる。

$$\sigma_2 = (i_2, \sigma(i_2), \dots, \sigma^{r_2-1}(i_2)) \quad (84)$$

この操作を続けると巡回置換 $\sigma_1, \sigma_2, \dots, \sigma_k$ が得られ、しかも σ は $\sigma = \sigma_1 \sigma_2 \dots \sigma_k$ と表せる。

一方で任意の巡回置換 $\tau = (i_1, i_2, \dots, i_r)$ は、互換 $(i_1, i_r), (i_1, i_{r-1}), \dots, (i_1, i_2)$ によって、

$$\tau = (i_1, i_r)(i_1, i_{r-1}) \dots (i_1, i_2) \quad (85)$$

と表される。以上により、任意の置換 σ は互換の積として表されることが分かる。

定理: 1 つの置換を互換の積で表すとき使われる互換の個数が偶数であるか奇数であるかは、その表し方によらないで定まる。

証明: 差積と呼ばれる以下の多項式を考える。

$$\begin{aligned} \Delta(x_1, \dots, x_n) &= \prod_{i < j} (x_i - x_j) = (x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_n) \\ &\quad (x_2 - x_3) \dots (x_2 - x_n) \\ &\quad \vdots \\ &\quad (x_{n-1} - x_n) \end{aligned} \quad (86)$$

Δ の変数 x_1, \dots, x_n を置換 σ によって、 $x_{\sigma(1)}, \dots, x_{\sigma(n)}$ に置き換えた式を $\sigma\Delta$ と書く。すなわち、

$$\sigma\Delta(x_1, \dots, x_n) = \Delta(x_{\sigma(1)}, \dots, x_{\sigma(n)}) \quad (87)$$

と書く。このとき、互換 $\tau = (i, j) (i < j)$ によって、

$$\tau\Delta(x_1, \dots, x_n) = -\Delta(x_1, \dots, x_n) \quad (88)$$

となることが分かる。実際、互換 $\tau = (i, j) (i < j)$ をほどこすと、

1. $x_i - x_j$ の項は $-(x_i - x_j)$
2. $(x_k - x_i)(x_k - x_j)$ の項は $(x_k - x_j)(x_k - x_i)$ ($k < i$)
3. $(x_i - x_k)(x_k - x_j)$ の項は $(x_j - x_k)(x_k - x_i)$ ($i < k < j$)
4. $(x_i - x_k)(x_j - x_k)$ の項は $(x_j - x_k)(x_i - x_k)$ ($j < k$)
5. それ以外の部分で不変

となる。最初の場合だけ符号が変化し、その他の場合は対になっている 2 つの因数の積は変わらない。今置換 σ が互換 $\tau_1, \dots, \tau_s, \rho_1, \dots, \rho_t$ を使って

$$\rho = \tau_1 \tau_2 \cdots \tau_k = \rho_1 \rho_2 \cdots \rho_t \quad (89)$$

の 2 通りに表せたとしても、

$$\begin{aligned} \sigma \Delta(x_1, x_2, \dots, x_n) &= (-1)^s \Delta(x_1, \dots, x_n) \\ &= (-1)^t \Delta(x_1, \dots, x_n) \end{aligned} \quad (90)$$

が得られる。従って、 s と t は共に奇数が偶数である。

置換 σ が偶数個の互換の積で表されるとき、 σ を偶置換、奇数個の積で表されるとき奇置換という。さらに置換 σ に対し、その符号を

$$\text{sgn } \sigma = \begin{cases} +1 & (\sigma \text{ が偶置換のとき}) \\ -1 & (\sigma \text{ が奇置換のとき}) \end{cases} \quad (91)$$

によって定義する。

問題 9-1: 置換 $\sigma = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{smallmatrix} \right)$ に対し、 $\text{sgn } \sigma$ を求めよ。

9.2 行列式の定義

n 次正方行列 $A = (a_{ij})$ に対し、以下の $|A|$ を A の行列式という。

$$|A| = \sum_{\sigma \in S_n} \text{sgn } \sigma \cdot a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \quad (92)$$

但し S_n は n 文字の置換全てを含む集合を表す。これは以下のようにも表される。

$$|A| = \det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \quad (93)$$

A が n 次正方行列のとき、 $|A|$ を n 次の行列式という。例として 2 次の行列式を求めてみよう。

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \text{sgn} \left(\begin{smallmatrix} 1 & 2 \\ 1 & 2 \end{smallmatrix} \right) a_{11} a_{22} + \text{sgn} \left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix} \right) a_{12} a_{21} = a_{11} a_{22} - a_{12} a_{21} \quad (94)$$

問題 9-2: 3 次の行列式 $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ を求めよ。

9.3 行列式の性質

ここではいくつか有用な行列式の性質について触れておこう。

定理: 正方行列 A の転置行列 A^t の行列式は A の行列式に等しい。すなわち、

$$|A^t| = |A| \quad (95)$$

証明: $A = (a_{ij})$ とし、 $A^t = (b_{ij})$ とすれば、 $b_{ij} = a_{ji}$ であるから、

$$|A^t| = \sum_{\sigma \in S_n} \text{sgn } \sigma b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)} = \sum_{\sigma \in S_n} \text{sgn } \sigma a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}$$

$\sigma^{-1} = \tau$ とおくと、

$$\sigma^{-1} = \left(\begin{array}{ccccc} \sigma(1) & \sigma(2) & \cdots & \sigma(n) \\ \downarrow & 1 & 2 & \cdots & n \end{array} \right) = \tau = \left(\begin{array}{ccccc} 1 & 2 & \cdots & n \\ \downarrow & \tau(1) & \tau(2) & \cdots & \tau(n) \end{array} \right)$$

ところで、 $\text{sgn } \sigma = \text{sgn } \sigma^{-1} = \text{sgn } \tau$ 。ここで $\{\sigma^{-1} | \sigma \in S_n\} = S_n$ であることに注意すれば、

$$\begin{aligned} |A^t| &= \sum_{\sigma \in S_n} \text{sgn } \sigma^{-1} a_{1\sigma^{-1}(1)} a_{2\sigma^{-1}(2)} \cdots a_{n\sigma^{-1}(n)} \\ &= \sum_{\tau \in S_n} \text{sgn } \tau a_{1\tau(1)} a_{2\tau(2)} \cdots a_{n\tau(n)} \\ &= |A| \end{aligned}$$

この定理により、行列式に関する性質で、行に関して成り立つことは全て列に関しても成り立ち、逆に列に関して成り立つことは全て行に対しても成り立つ (行と列の双対性)。

定理: 行列式において、ある行を c 倍したものは全体を c 倍したものに等しい。すなわち、

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & & & \vdots \\ ca_{i1} & ca_{i2} & \cdots & \cdots & ca_{in} \\ \vdots & \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{vmatrix} = c \begin{vmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & a_{in} \\ \vdots & \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{vmatrix} \quad (96)$$

定理: 行列式において、ある行の c 倍を別の行に足したものはもとの行列式に等しい。すなわち、

$$\begin{vmatrix}
a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\
\vdots & \vdots & & & \vdots \\
a_{i1} & a_{i2} & \cdots & \cdots & a_{in} \\
\vdots & \vdots & & & \vdots \\
a_{k1} + ca_{i1} & a_{k2} + ca_{i2} & \cdots & \cdots & a_{kn} + ca_{in} \\
\vdots & \vdots & & & \vdots \\
a_{n1} & a_{n2} & \cdots & \cdots & a_{nn}
\end{vmatrix} = \begin{vmatrix}
a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\
\vdots & \vdots & & & \vdots \\
a_{i1} & a_{i2} & \cdots & \cdots & a_{in} \\
\vdots & \vdots & & & \vdots \\
a_{k1} & a_{k2} & \cdots & \cdots & a_{kn} \\
\vdots & \vdots & & & \vdots \\
a_{n1} & a_{n2} & \cdots & \cdots & a_{nn}
\end{vmatrix} \quad (97)$$

定理: 先頭を除いて一行目の要素が 0 の行列式は先頭の要素と先頭の行・列を削除した行列式との積に等しい。すなわち、

$$\begin{vmatrix}
a_{11} & 0 & \cdots & 0 \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{vmatrix} = a_{11} \begin{vmatrix}
a_{22} & \cdots & a_{2n} \\
\vdots & & \vdots \\
a_{n2} & \cdots & a_{nn}
\end{vmatrix} \quad (98)$$

問題 9-3: $\begin{vmatrix} -1 & 0 & 1 & 1 \\ 2 & -1 & 0 & 2 \\ 1 & 2 & 1 & -1 \\ -1 & -1 & 1 & 0 \end{vmatrix}$ の値を求めよ。

9.4 内積

任意のベクトル a, b に対し、実数 $a \cdot b$ が対応し、 $a \cdot b$ が以下の条件を全て満たすとき、 $a \cdot b$ を a と b の内積という。

1. $a \cdot b = b \cdot a$
2. $(\lambda a) \cdot b = \lambda(a \cdot b)$
3. $(a + b) \cdot c = a \cdot c + b \cdot c$
4. 常に $a \cdot a \geq 0$ 。但し等号が成り立つのは、 $a = o$ のときに限る (必要十分条件)。

任意のベクトル v に対し、 $\|v\| = \sqrt{v \cdot v}$ を v の長さ、またはノルムという。 $a \cdot b = \|a\| \|b\| \cos \theta$ (但し θ はベクトル a と b のなす角) とすれば、これは上の条件を全て満たすことを確かめられたい。従ってこの定義による $a \cdot b$ は a と b の内積と呼べる。

$a \cdot b = 0$ のとき、 a と b は直交するという。 o でないベクトル a_1, a_2, \dots, a_n はどの 2 つも直交するとき、直交系と呼ばれる。さらに単位ベクトルからなる直交系、すなわち、

$$u_i \cdot u_j = \delta_{ij}$$

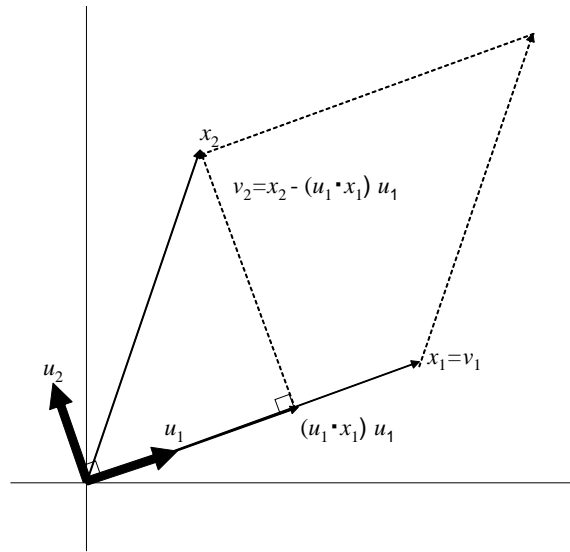


図 12: 二次元の場合のベクトル直交化

を満たすベクトル u_1, u_2, \dots, u_n を正規直交系と呼ぶ。ここで δ_{ij} は

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

と定義され、クロネッカー (Kronecker) のデルタと呼ばれる。

問題 9-4: 2 つのベクトル $a = (a_1, a_2, \dots, a_n)^t$, $b = (b_1, b_2, \dots, b_n)^t$ の内積が $a \cdot b = \|a\| \|b\| \cos \theta$ (但し θ はベクトル a と b のなす角) で定義されるとき、余弦定理⁸ を使って

$$a \cdot b = \sum_{i=1}^n a_i b_i \quad (99)$$

を示せ。

9.5 グラム・シュミットの直交化法

線形独立なベクトル x_1, x_2, \dots, x_n に対し、 u_1, u_2, \dots, u_n を順次以下の手順で定めれば、 u_1, u_2, \dots, u_n は正規直交系である。

⁸三角形の内角を A, B, C 、これと向き合う辺の長さをそれぞれ α, β, γ とすると、 $\gamma^2 = \alpha^2 + \beta^2 - 2\alpha\beta \cos C$ が成立する。

$$\begin{aligned}
v_1 &= x_1 & u_1 &= \frac{v_1}{\|v_1\|} \\
v_2 &= x_2 - (u_1 \cdot x_2)u_1 & u_2 &= \frac{v_2}{\|v_2\|} \\
v_3 &= x_3 - (u_1 \cdot x_3)u_1 - (u_2 \cdot x_3)u_2 & u_3 &= \frac{v_3}{\|v_3\|} \\
\vdots & & \vdots & \\
v_n &= x_n - \sum_{k=1}^{n-1} (u_k \cdot x_n)u_k & u_n &= \frac{v_n}{\|v_n\|}
\end{aligned}$$

これをグラム シュミット (Gram-Schmidt) の直交化法という。

問題 9-5: 次のベクトルからグラム シュミットの直交化法により正規直交系を作れ。

$$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

9.6 行列式と超平行面体の超体積との関係

図 12 の v_2 に着目しよう。 v_2 は x_2 から「 x_2 の u_1 方向に平行な成分」を引いたベクトル、すなわち平行四辺形の面積を「底辺の長さ×高さ」として u_1 の絶対値を底辺の長さとする、 v_2 の長さは高さに相当する。故に、 x_1 と x_2 で構成される平行四辺形の面積 (底辺の長さ×高さ) は、 $v_1 = x_1$ と v_2 で構成される長方形の面積と等しい。

同様の議論を繰り返していくと、 $(x_1 \ x_2 \ \cdots \ x_n)$ で構成される平行 $2n$ 面体の体積は、 $(v_1 \ v_2 \ \cdots \ v_n)$ で構成される超直方体の体積と等しくなることが分かる。超直方体の体積 $\|V\|$ は、

$$\|V\| = \prod_{k=1}^n \|v_k\| \tag{100}$$

になる。

さて、 $(x_1 \ x_2 \ \cdots \ x_n)$ の行列式をグラム シュミットの方法と絡めながら求めよう。式 96 と式 97 に注意すれば、

$$\begin{aligned}
& |x_1 \ x_2 \ \cdots \ x_n| \\
&= \|x_1\| \left| \frac{x_1}{\|x_1\|} \ x_2 \ \cdots \ x_n \right| \\
&= \|x_1\| |u_1 \ x_2 \ \cdots \ x_n| \\
&= \|x_1\| |u_1 \ x_2 - (u_1 \cdot x_2)u_1 \ \cdots \ x_n| \\
&= \|x_1\| \|x_2 - (u_1 \cdot x_2)u_1\| \left| u_1 \ \frac{x_2 - (u_1 \cdot x_2)u_1}{\|x_2 - (u_1 \cdot x_2)u_1\|} \ \cdots \ x_n \right| \\
&= \|x_1\| \|x_2 - (u_1 \cdot x_2)u_1\| |u_1 \ u_2 \ \cdots \ x_n|
\end{aligned}$$

$$\begin{aligned}
& \vdots \\
& = \|x_1\| \|x_2 - (u_1 \cdot x_2)\| \cdots \|x_n - \sum_{k=1}^{n-1} (u_k \cdot x_n) u_k\| |u_1 \ u_2 \ \cdots \ u_n| \\
& = |u_1 \ u_2 \ \cdots \ u_n| \prod_{k=1}^n \|v_k\| = |u_1 \ u_2 \ \cdots \ u_n| \|V\|
\end{aligned} \tag{101}$$

を得る。正規直交ベクトルで構成された行列 $U = (u_1 \ u_2 \ \cdots \ u_n)$ は直交行列 $U^t U = E$ を成し、 $|U| = \pm 1$ である。よって $\| |x_1 \ x_2 \ \cdots \ x_n| \| = \|V\|$ を得る。すなわち、 $(x_1 \ x_2 \ \cdots \ x_n)$ の行列式の絶対値は、これが成す平行 $2n$ 面体の体積と等しい。

問題 9-6: 2 つのベクトル $a = (a_1, a_2)^t$ と $b = (b_1, b_2)^t$ が形成する平行四辺形の面積を求めよ。

10 多変量の確率関数

10.1 多変量の確率関数の定義

確率関数は多変量に対しても拡張できる。まずは 2 変量で考えてみよう。確率密度関数を $f(x, y)$ とすれば、節 1.3 で述べた確率密度関数の性質を以下のように拡張できる。

1. 確率変数を取りうる全範囲の確率の合計は 1 となる。すなわち、

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1 \tag{102}$$

2. X, Y が範囲 D の値に収まる確率は以下の式で与えられる。

$$P((X, Y) \in D) = \iint_D f(x, y) dx dy \tag{103}$$

図形的に表せば図 13 のように、 X, Y が領域 D に入る確率は領域 D と $f(x, y)$ の部分で囲まれた空間の体積で表される。すなわち、

$$P((x_i \leq X < x_i + \Delta x) \wedge (y_j \leq Y < y_j + \Delta y)) \approx f(x_i, y_j) \Delta x \Delta y \tag{104}$$

領域 D に $\Delta x, \Delta y$ の長方形を互いに重ならないようにタイルのように可能な限り敷き詰め、 i 番目の長方形の隅の座標を $x_i, y_j \in D$ として合計すると、 $\sum_{(x_i, y_j) \in D} f(x_i, y_j) \Delta x \Delta y$ であり、 $\Delta x, \Delta y \rightarrow 0$ のとき、式 103 に収束する。

これをさらに n 変量で考えると、

1. 確率変数を取りうる全範囲の確率の合計は 1 となる。すなわち、

$$\int_{x_1=-\infty}^{x_1=+\infty} \int_{x_2=-\infty}^{x_2=+\infty} \cdots \int_{x_n=-\infty}^{x_n=+\infty} f(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n = 1 \tag{105}$$

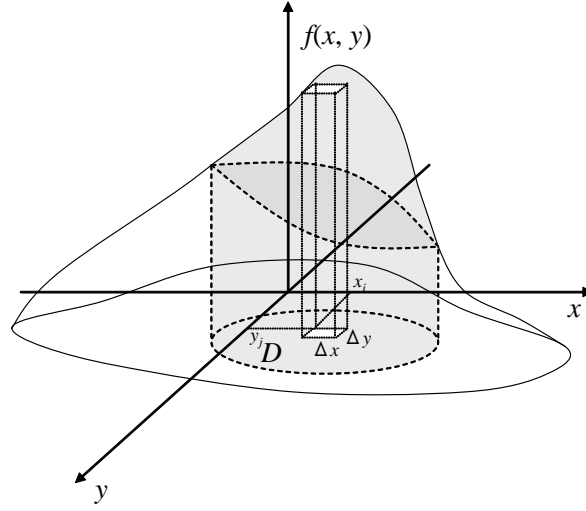


図 13: 3次元空間上の連続分布

2. X_1, X_2, \dots, X_n が範囲 D の値に収まる確率は以下の式で与えられる。

$$P((X_1, X_2, \dots, X_n) \in D) = \int \int \cdots \int_D f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n \quad (106)$$

式 105、106 のいずれにおいても積分記号 \int は n 回登場することに注意されたい。

確率密度関数 $X \sim f(x), Y \sim g(y)$ について、

$$\begin{aligned} P(x \leq X \leq x + \Delta x) &\approx f(x) \Delta x \\ P(y \leq Y \leq y + \Delta y) &\approx g(y) \Delta y \end{aligned} \quad (107)$$

が成立するが、 X と Y が独立な場合、

$$P((x \leq X \leq x + \Delta x) \wedge (y \leq Y \leq y + \Delta y)) = f(x) \Delta x \cdot g(y) \Delta y = f(x)g(y) \Delta x \Delta y \quad (108)$$

となり、式 104 で $f(x, y) = f(x)g(y)$ としたもの的一致する。すなわち、 $f(x)g(y)$ は X, Y が従う結合密度関数である。

確率変数 X と Y の和 $X + Y$ が従う分布 $h(z)$ は以下のように求めることができる。まず $X \sim f(x), Y \sim g(y)$ とすると、その結合密度関数は、 $f(x)g(y)$ である。 $a \leq X + Y \leq z$ となる確率は、

$$H(z) = \int_{-\infty}^z h(w) dw \quad (109)$$

とおくと、

$$P(a \leq X + Y \leq z) = H(z) - H(a) \quad (110)$$

と表すことができる。次に $a \leq X + Y \leq z$ を満たす X と Y の範囲を考えよう。 Y が $-\infty$ から $+\infty$ まで自由に値を変えるとすると、 X は $a - Y \leq X \leq z - Y$ という制約を受ける。この制約の中で $f(x)g(y)$ を積分すれば、 $P(a \leq X + Y \leq z)$ が求まる。すなわち $F(z) = \int_{-\infty}^z f(x)dx$ とおけば、

$$\begin{aligned} P(a \leq X + Y \leq z) &= H(z) - H(a) = \int_{y=-\infty}^{y=+\infty} \left\{ \int_{x=a-y}^{x=z-y} f(x)dx \right\} g(y)dy \\ &= \int_{y=-\infty}^{y=+\infty} \{F(z-y) - F(a-y)\} g(y)dy \\ &= \int_{y=-\infty}^{y=+\infty} F(z-y)g(y)dy - \int_{y=-\infty}^{y=+\infty} F(a-y)g(y)dy \end{aligned} \quad (111)$$

z を含む項、 a を含む項だけを抜き出すと、

$$H(z) = \int_{y=-\infty}^{y=+\infty} F(z-y)g(y)dy, \quad H(a) = \int_{y=-\infty}^{y=+\infty} F(a-y)g(y)dy \quad (112)$$

左の式について z で微分すると、

$$h(z) = \int_{-\infty}^{+\infty} f(z-y)g(y)dy \quad (113)$$

が求まる。式 113 を f と g の畳み込み積分 (convolution) という。

畳み込み積分を用いて $N(\mu, \sigma^2)$ に従う 2 つの確率変数 X, Y の和 $X + Y$ が従う分布を求めよう。 $N(\mu, \sigma^2)$ の確率密度関数を $f(x)$ として、 $X + Y$ の従う結合密度関数は、

$$\begin{aligned} h(z) &= \int_{-\infty}^{+\infty} f(z-y)f(y)dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-y-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\frac{1}{2}(z-2\mu)^2}{2\sigma^2}} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{2(y-\frac{1}{2}z)^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-2\mu)^2}{4\sigma^2}} \cdot \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{2}}\sigma} e^{-\frac{(y-\frac{1}{2}z)^2}{2(\frac{1}{\sqrt{2}}\sigma)^2}} dy \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} e^{-\frac{(z-2\mu)^2}{2(\sqrt{2}\sigma)^2}} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{2}}\sigma} e^{-\frac{(y-\frac{1}{2}z)^2}{2(\frac{1}{\sqrt{2}}\sigma)^2}} dy \\ &= N(2\mu, 2\sigma^2) \text{ の確率密度関数} \cdot N\left(\frac{1}{2}z, \frac{1}{2}\sigma^2\right) \text{ の全域に関する積分} = N(2\mu, 2\sigma^2) \text{ の確率密度関数} \cdot 1 \\ &= N(2\mu, 2\sigma^2) \text{ の確率密度関数} \quad (114) \end{aligned}$$

となり、 $X + Y$ は $N(2\mu, 2\sigma^2)$ に従うことが示された。

問題 10-1: 畳み込み積分を用いて、サイコロルーレットを 2 回回したときに得られる数値の合計値が従う確率密度関数を求めよ。

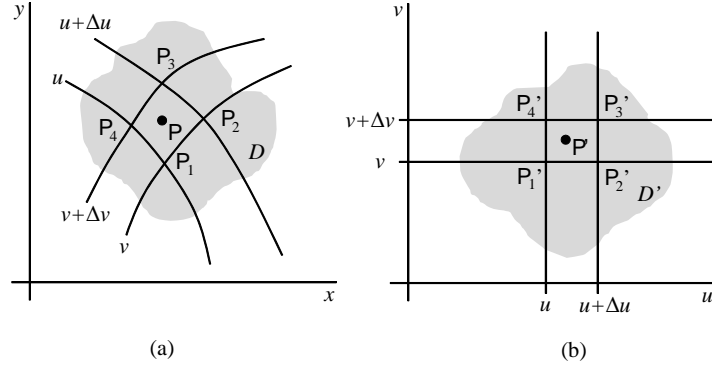


図 14: 2 変数の変換のイメージ

10.2 多変量の確率関数に関する期待値計算

確率変数 X が確率密度関数 $f(x)$ に従うとする。また同時に X は n 個の変数で定義される関数 u で $X = u(X_1, X_2, \dots, X_n)$ と表され、さらに X_1, X_2, \dots, X_n は確率密度関数 $g(x_1, x_2, \dots, x_n)$ に従うとする。また関数 h が与えられているとき、 $E(h(X))$ を求めてみよう。

まず、 $E(h(X))$ は $h(x) \times P(x \leq X < x + \Delta x)$ の確率の総和であるという観点から、

$$E(h(X)) = \int_{-\infty}^{+\infty} h(x)f(x)dx \quad (115)$$

である。また確率密度関数に φ を使って考えると、 $E(h(u(x_1, x_2, \dots, x_n)))$ は $h(u(x_1, x_2, \dots, x_n)) \times P((x_1, x_2, \dots, x_n) \in \Delta D)$ の総和と考えれば、

$$E(h(X)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(u(x_1, x_2, \dots, x_n))g(x_1, x_2, \dots, x_n)dx_1dx_2 \cdots dx_n \quad (116)$$

となる。

10.3 多変量の確率関数の変数変換

$x = \varphi(u, v), y = \psi(u, v)$ の関係が成立するとき、図 14 に示すように、 (x, y) の定義域 D の 1 小領域 $P_1P_2P_3P_4$ とし、その内部の点 P がある。上式によって $P_1P_2P_3P_4$ が $P'_1P'_2P'_3P'_4$ に移ったとする。点 P_i を (x_i, y_i) で表せば、

$$\lim_{\Delta u \rightarrow 0} \frac{\varphi(u + \Delta u, v) - \varphi(u, v)}{\Delta u} = \frac{\partial \varphi}{\partial u}$$

$$\lim_{\Delta u \rightarrow 0} \lim_{\Delta v \rightarrow 0} \varphi(u + \Delta u, v + \Delta v) - \varphi(u, v) = \frac{\partial \varphi}{\partial u} \Delta u + \frac{\partial \varphi}{\partial v} \Delta v \quad (\text{全微分})$$

などに注意して、

$$\begin{aligned}
x_1 &= \varphi(u, v) \\
y_1 &= \psi(u, v) \\
x_2 &= \varphi(u + \Delta u, v) = \varphi(u, v) + \frac{\partial \varphi}{\partial u} \Delta u \\
y_2 &= \psi(u + \Delta u, v) = \psi(u, v) + \frac{\partial \psi}{\partial u} \Delta u \\
x_3 &= \varphi(u + \Delta u, v + \Delta v) = \varphi(u, v) + \frac{\partial \varphi}{\partial u} \Delta u + \frac{\partial \varphi}{\partial v} \Delta v \\
y_3 &= \psi(u + \Delta u, v + \Delta v) = \psi(u, v) + \frac{\partial \psi}{\partial u} \Delta u + \frac{\partial \psi}{\partial v} \Delta v \\
x_4 &= \varphi(u + v + \Delta v) = \varphi(u, v) + \frac{\partial \varphi}{\partial v} \Delta v \\
y_4 &= \psi(u + v + \Delta v) = \psi(u, v) + \frac{\partial \psi}{\partial v} \Delta v
\end{aligned} \tag{117}$$

平行四辺形の面積 ΔS は、

$$\left| \begin{array}{cc} \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial v} \\ \frac{\partial \psi}{\partial u} & \frac{\partial \psi}{\partial v} \end{array} \right| \Delta u \Delta v = \frac{\partial(x, y)}{\partial(u, v)} \Delta u \Delta v = J \Delta u \Delta v \tag{118}$$

J はヤコビアン (Jacobian) と呼ばれる。 $\Delta S' = \Delta u \Delta v$ として、

$$\Delta S = |J| \Delta S' \tag{119}$$

を得る。 D 上で定義された $f(x, y)$ の x, y に最初の式を代入すれば、 $f(\varphi(u, v), \psi(u, v))$ は D' 上の関数である。 D' 上の分割による小領域 $\Delta S'_i$ に対し、 D 上の小領域 ΔS が対応するとすれば、

$$\sum_{i=1}^n f(x_i, y_i) \Delta S_i = \sum_{i=1}^n f(\varphi(u, v), \psi(u, v)) |J| \Delta S'_i \tag{120}$$

各 $\Delta S'_i \rightarrow 0$ となる極限を考えれば、 x, y は連続であるから、 $\Delta S_i \rightarrow 0$ となり、

$$\int \int_D f(x, y) dx dy = \int \int_{D'} f(\varphi(u, v), \psi(u, v)) |J| du dv \tag{121}$$

11 積率母関数

11.1 積率母関数の定義

x を確率変数とし、 t を実数とすると、 e^{tx} の期待値を x の積率母関数 (moment generating function) といい、 $M_x(t)$ で表す。 すなわち x が離散型であれば、 $f(x)$ を確率関数として、

$$M_x(t) = E(e^{tx}) = \sum_x e^{tx} f(x) \tag{122}$$

となる。但し、 σ_x は x の可能な値全てについての合計を表す。また x が連続型で $-\infty < x < \infty$ とすれば、 $f(x)$ を密度関数として、

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad (123)$$

である。

e^{tx} をマクローリン展開すれば、

$$e^{tx} = 1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \cdots \quad (124)$$

であるから、これを式 123 に代入して、

$$\begin{aligned} E(e^{tx}) &= \int_{-\infty}^{\infty} f(x) dx + \int_{-\infty}^{\infty} tx f(x) dx + \frac{1}{2} \int_{-\infty}^{\infty} (tx)^2 f(x) dx + \cdots \\ &= 1 + tE(x) + \frac{t^2}{2} E(x^2) + \cdots = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(x^k) \end{aligned} \quad (125)$$

となる。すなわち各項は原点回りの k 次の積率 $E(x^k)$ を含んでいる。

式 125 に関する k 階の導関数で $t = 0$ とおくとその値は $E(x^k)$ すなわち原点回りの k の積率に等しくなる。すなわち、

$$\left. \frac{\partial^k M_x(t)}{\partial t^k} \right|_{t=0} = E(x^k) \quad k = 1, 2, \cdots \quad (126)$$

これが $M_x(t)$ が積率母関数と呼ばれる所以である。以下、 $\partial^k M_x(t) / \partial t^k|_{t=0}$ を単に $M_x^{(k)}(0)$ と表すことにする。例えば $k = 1$ であれば、

$$M'_x(0) = E(x)$$

で平均が得られる。また $k = 2$ ならば、

$$M''_x(0) = E(x^2)$$

すなわち、原点周りの 2 次の積率が得られるから、

$$E(x^2) - E(x)^2 = \sigma^2$$

によって分散が得られる。

例題 11-1: 正規分布 $N(\mu, \sigma^2)$ の積率母関数を求めてみよう。

$$\begin{aligned} M_x(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma^2} \underbrace{\{(x-\mu)^2 - 2\sigma^2 tx\}}_* \right] dx \end{aligned}$$

ところで * の部分は $(x - \mu)^2 - 2\sigma^2 tx = x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx = x^2 - 2(\mu - \sigma^2 t)x + \mu^2 = \{x - (\mu + \sigma^2 t)\}^2 - 2\sigma^2 \mu t - \sigma^4 t^2$ であるから、

$$\begin{aligned} M_x(t) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma^2} \{x - (\mu + \sigma^2 t)\}^2 + \mu t + \frac{\sigma^2}{2} t^2 \right] dx \\ &= \exp \left(\mu t + \frac{\sigma^2}{2} t^2 \right) \cdot \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma^2} \{x - (\mu + \sigma^2 t)\}^2 \right] dx \end{aligned}$$

ここで、 $\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma^2} \{x - (\mu + \sigma^2 t)\}^2 \right] dx$ は平均 $\mu + \sigma^2 t$ 、分散 σ^2 の正規分布密度関数の積分に他ならないから、1 に等しい。それゆえ、

$$M_x(t) = \exp \left(\mu t + \frac{\sigma^2}{2} t^2 \right) \quad (127)$$

である。

$M'_x(t) = (\mu + \sigma^2 t) \exp(\mu t + \frac{\sigma^2}{2} t^2)$ であるから、

$$M'_x(0) = \mu \quad (128)$$

となり、また $M''_x(t) = \sigma^2 \exp(\mu t + \frac{\sigma^2}{2} t^2) + (\mu + \sigma^2 t)^2 \exp(\mu t + \frac{\sigma^2}{2} t^2)$ から、 $M''(0) = \sigma^2 + \mu^2$ 、すなわち、

$$M''(0) - M'(0)^2 = \sigma^2 \quad (129)$$

となり、正規分布の平均は μ 、分散は σ^2 であることが示された。

11.2 確率変数の関数の積率母関数

x を確率変数、 t を実数とすると、 e^{tx} の期待値を x の積率母関数ということは既に述べた。ここで、 $\varphi(x)$ を x の連続な関数としたとき、 $\varphi(x)$ の積率母関数は $e^{t\varphi(x)}$ の期待値と定義され、 $M_{\varphi(x)}(t)$ で表される。

x が連続型で $-\infty < x < \infty$ とすれば、 $f(x)$ を密度関数として、

$$M_{\varphi(x)}(t) = E(e^{t\varphi(x)}) = \int_{-\infty}^{\infty} e^{t\varphi(x)} f(x) dx \quad (130)$$

である。

$e^{t\varphi(x)}$ をマクローリン展開すれば、

$$e^{t\varphi(x)} = 1 + \frac{t\varphi(x)}{1!} + \frac{(t\varphi(x))^2}{2!} + \cdots \quad (131)$$

であるから、これを式 130 に代入して、

$$\begin{aligned}
E(e^{t\varphi(x)}) &= \int_{-\infty}^{\infty} f(x)dx + \int_{-\infty}^{\infty} t\varphi(x)f(x)dx + \frac{1}{2} \int_{-\infty}^{\infty} \{t\varphi(x)\}^2 f(x)dx + \cdots \\
&= 1 + tE[\varphi(x)] + \frac{t^2}{2}E[\{\varphi(x)\}^2] + \cdots
\end{aligned} \tag{132}$$

であるから、

$$M^k(0) = E[\{\varphi(x)\}^k] \quad k = 1, 2, \dots \tag{133}$$

により、 $\varphi(x)$ の原点回りの k 次の積率が計算できる。

11.3 正規分布による二項分布の近似

a, b および p を一定とすると、確率

$$P(a \leq x \leq b) = \sum_{a \leq x \leq b} \frac{n!}{x!(n-x)!} p^x q^{n-x}, \quad q = 1 - p \tag{134}$$

は $n \rightarrow \infty$ のとき、

$$\int_{(a-np)/\sqrt{npq}}^{(b-np)/\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \tag{135}$$

に収束する。

証明：標準化された $z = (x - np)/\sqrt{npq}$ の積率母関数は、

$$\begin{aligned}
M_z(t) &= \sum_{x=0}^n \frac{n!}{x!(n-x)!} p^x q^{n-x} e^{t(x-np)/\sqrt{npq}} = e^{-\frac{np}{\sqrt{npq}}t} \sum_{x=0}^n \frac{n!}{x!(n-x)!} (pe^{\frac{t}{\sqrt{npq}}})^x q^{n-x} \\
&= e^{-\frac{np}{\sqrt{npq}}t} (q + pe^{\frac{t}{\sqrt{npq}}})^n
\end{aligned} \tag{136}$$

従って、

$$\log M_z(t) = -\frac{np}{\sqrt{npq}}t + n \log(q + pe^{\frac{t}{\sqrt{npq}}}) = -\frac{np}{\sqrt{npq}}t + n \log(1 + u) \tag{137}$$

但し $u = p(-1 + e^{t/\sqrt{npq}})$ となる。 $\log(1 + u)$ を展開すれば、

$$\log(1 + u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \cdots \tag{138}$$

となる。また u は $e^{t/\sqrt{npq}}$ の展開により、

$$u = p(-1 + e^{\frac{t}{\sqrt{npq}}}) = p \left\{ \left(\frac{t}{\sqrt{npq}} \right) + \frac{1}{2!} \left(\frac{t}{\sqrt{npq}} \right)^2 + \frac{1}{3!} \left(\frac{t}{\sqrt{npq}} \right)^3 + \cdots \right\} \tag{139}$$

となる。式 139 を式 138 に代入し、これをさらに式 137 に代入して整理すれば、

$$\log M_z(t) = \frac{t^2}{2} + \sum_{k=3}^{\infty} c_k \left(\frac{t}{\sqrt{npq}} \right)^k \quad (140)$$

となる。但し右辺の第 2 項は $\frac{t}{\sqrt{npq}}$ の 3 次以上の項よりなる多項式で、 c_3, c_4, \dots は n を含まない係数である。そこで両辺の極限を取れば $\lim_{n \rightarrow \infty} \log M_x(t) = \frac{t^2}{2}$ 、すなわち

$$\lim_{n \rightarrow \infty} M_z(t) = e^{\frac{t^2}{2}} \quad (141)$$

となる。これは正規分布 $N(0, 1)$ の積率母関数に他ならない。よって、 $z = (x - np)/\sqrt{npq}$ は $N(0, 1)$ に分布収束する。このことは、 $P(a \leq x \leq b) = P((a - np)/\sqrt{npq} \leq z \leq (b - np)/\sqrt{npq})$ が式 135 に収束することを意味する。

12 中心極限定理

定理: x_1, \dots, x_n を密度関数 $f(x)$ からの大きさ n の無作為標本とし、 x の平均を μ 、分散を σ^2 とする。またこの標本の標本平均を \bar{x} 、標本平均を標準化した変数を z_* とする。すなわち、

$$z_* = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = z\sqrt{n} \quad (142)$$

このとき $z_* \sim N(0, 1)$ 、すなわち z_* は標準正規分布に従う。この定理を中心極限定理という。

証明: z_* の積率母関数は、

$$\begin{aligned} M_z(t) &= \int_{-\infty}^{\infty} e^{tz_*} f(z_*) dz_* \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} t\right) f(x_1) \cdots f(x_n) dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(t \sum_{i=1}^n \frac{x_i - \mu}{\sigma/\sqrt{n}}\right) f(x_1) \cdots f(x_n) dx_1 \cdots dx_n \\ &= \left\{ \int_{-\infty}^{\infty} \exp\left(\frac{x - \mu}{\sigma/\sqrt{n}} t\right) f(x) dx \right\}^n = \left\{ M_z\left(\frac{t}{\sqrt{n}}\right) \right\}^n \\ &= \left\{ \int_{-\infty}^{\infty} \exp\left(\frac{t}{\sqrt{n}} z\right) f(z) dz \right\}^n \end{aligned} \quad (143)$$

ここで、

$$\exp\left(\frac{t}{\sqrt{n}} z\right) = 1 + \frac{tz}{\sqrt{n}} + \frac{1}{2} \left(\frac{tz}{\sqrt{n}}\right)^2 + \frac{1}{6} \left(\frac{tz}{\sqrt{n}}\right)^3 + \cdots \quad (144)$$

となるから、

$$\begin{aligned} \left\{ M_z \left(\frac{t}{\sqrt{n}} \right) \right\}^n &= \left[\int_{-\infty}^{\infty} \left\{ 1 + \frac{tz}{\sqrt{n}} + \frac{1}{2} \left(\frac{tz}{\sqrt{n}} \right)^2 + \frac{1}{6} \left(\frac{tz}{\sqrt{n}} \right)^3 + \cdots \right\} f(z) dz \right]^n \\ &= \left[\int_{-\infty}^{\infty} f(z) dz + \int_{-\infty}^{\infty} \frac{t}{\sqrt{n}} z f(z) dz + \int_{-\infty}^{\infty} \frac{t^2}{2n} z^2 f(z) dz + \int_{-\infty}^{\infty} \frac{t^3}{6n^{3/2}} z^3 f(z) dz + \cdots \right]^n \end{aligned} \quad (145)$$

となる。ここで、 $\int_{-\infty}^{\infty} f(z) dz = 1$ (確率密度関数の定義、全確率の合計は1)、 $\int_{-\infty}^{\infty} z f(z) dz = 0$ (z の平均は0)、 $\int_{-\infty}^{\infty} z^2 f(z) dz = 1$ (z の分散) より、式 145 は、

$$M_{z*}(t) = \left[1 + 0 + \frac{t^2}{2n} + (n^{-3/2} \text{の項}) + (n^{-2} \text{の項}) + \cdots \right]^n \quad (146)$$

かくして、両辺の $n \rightarrow \infty$ における極限を取れば

$$\lim_{n \rightarrow \infty} M_{z*}(t) = e^{\frac{t^2}{2}} \quad (147)$$

となり、標準正規分布の積率母関数と一致する。

問題 12-1: 式 147 が成り立つことを示せ。

略解

3-1: $E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$

3-2: 3-1 の結果および $E(X^2) = \sigma^2 + \mu^2$ を利用して、

$$\begin{aligned} E\left\{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right\} &= \frac{1}{n} \sum E\{X_i^2 - 2X_i\bar{X} + \bar{X}^2\} = \frac{1}{n} \sum E(X_i^2) - \frac{2}{n} E(\sum X_i \bar{X}) + \frac{1}{n} E(\sum \bar{X}^2) \\ &= \frac{1}{n} n(\mu^2 + \sigma^2) - \frac{2}{n} E(n\bar{X}\bar{X}) + \frac{1}{n} E(n\bar{X}^2) = \mu^2 + \sigma^2 - E(\bar{X}^2) = \mu^2 + \sigma^2 - \frac{\sigma^2}{n} - \mu^2 = \left(1 - \frac{1}{n}\right) \sigma^2 \end{aligned}$$

5-1: $y = \tan x$ として、 $\frac{dy}{dx} = \frac{1}{\cos^2 x}$ 。従って、 $\frac{dx}{dy} = \frac{d \tan^{-1} y}{dy} = \cos^2 x = \frac{1}{\tan^2 x + 1} = \frac{1}{y^2 + 1}$ 。

10-1: サイコロルーレットを 1 回回したときに得られる数値 x に関する確率密度関数 $f(x)$ は、

$$f(x) = \begin{cases} \frac{1}{6} & \text{for } 0 \leq x \leq 6 \\ 0 & \text{otherwise} \end{cases} \quad \text{であり、2 回回したときに得られる数値の合計値 } z \text{ に関する確率密}$$

度関数は畳み込み積分を行って、 $h(z) = \int_{-\infty}^{+\infty} f(z-y)g(y)dy$ である。 y に関する積分を行う区間は $f(z-y)g(y) > 0$ 、すなわち、 $f(z-y) > 0$ かつ $g(y) > 0$ の範囲である。これを満たす y の範囲は、 $0 \leq z-y \leq 6$ かつ $0 \leq y \leq 6$ 、すなわち $z-6 \leq y \leq z$ かつ $0 \leq y \leq 6$ である。 $z < 0$ または $z > 12$ のときはこの制約を満たすことができず、 $h(z) = 0$ となる。 $0 \leq z \leq 6$ のときは、 $0 \leq y < z$ 、 $6 \leq z \leq 12$ のときは、 $z-6 \leq y \leq 6$ が積分する範囲となる。すなわち、

$$h(z) = \begin{cases} \frac{1}{36}z & \text{for } 0 \leq z \leq 6 \\ -\frac{1}{36}z + \frac{1}{3} & \text{for } 6 \leq z \leq 12 \\ 0 & \text{otherwise} \end{cases}$$

12-1: $\lim_{n \rightarrow \infty} n \cdot h(n) = 0$ なら、任意に小さい $\epsilon > 0$ に対し、 $\exists_{n_0} \forall_{n > n_0} \{0 < n \cdot h(n) < \epsilon\}$ が成立するから、十分大きな n に対し、 $\frac{1}{n} < \frac{1}{n} + h(n) < \frac{1+\epsilon}{n}$ とすることができる。 $(1 + \frac{1}{n})^n < (1 + \frac{1}{n} + h(n))^n < (1 + \frac{1+\epsilon}{n})^n$ より、 $n \rightarrow \infty$ をとって、 $e < \lim_{n \rightarrow \infty} (1 + \frac{1}{n} + h(n))^n < e^{1+\epsilon}$ 。さらに $n \rightarrow \infty$ のとき、 $e^{1+\epsilon}$ はいくらでも e に近づけるから、 $\lim_{n \rightarrow \infty} (1 + \frac{1}{n} + h(n))^n = e$ 。

参考文献

- [1] 富田勝 (監修) 斎藤輪太郎 (著) 「バイオインフォマティクスの基礎～ゲノム解析プログラミングを中心に～」サイエンス社 (2005/07)
- [2] Hall, P(1927) "The Distribution of Means for Samples of Size N Drawn from a Population in which the Variate Takes Values Between 0 and 1, All Such Values Being Equally Probable". Biometrika, Vol. 19, No. 3/4., pp. 240-245.
- [3] Irwin JO (1927) "On the Frequency Distribution of the Means of Samples from a Population Having any Law of Frequency with Finite Moments, with Special Reference to Pearson's Type II". Biometrika, Vol. 19, No. 3/4., pp. 225-239.
- [4] 黒川巖 (発行) 金井康彦 (編集) 「学研の図鑑 数・形」学習研究社 (1977)