正規分布

1　正規分布の発見

どんなに偏りのあるコインであっても，繰り返し投げ続けていくと，すべてが同じ結果，たとえば表ばかりとか裏ばかりではなく，ある変動をもった結果が得られると想像される。ｎ枚のコインを投げると（１枚のコイン投げをｎ回繰り返すと），表の出る枚数（回数）は２項分布にしたがうことという命題は既に学んだ。この枚数（回数）をどんどん大きくしていくと，２項分布ではあるが，ある分布に近づくことがわかる。この分布が正規分布である。ラプラス (Pierre Simon de Laplace), 1749-1827 フランスの革命期の数学者、天文学者。数学特に解析学の多くの分野に大きな業績を残したが、古典確率論の大成はその貢献の一つである。一方，ガウス (Karl Friedrich Gauss), 1777-1855　ドイツの数学者、物理学者、天文学者。幼少時より天才の誉れ高く、数学・物理学に巨大な功績を残しているが，この分野では正規分布の発見がある。正規分布を２項分布の極限として，ラプラスが導いた方法とは全く異なり，天体観測の誤差測定を解析し，変動があり得る状況から，微分方程式をたてて，正規分布を導いた。「誤差論」円錐曲線で太陽の回りを回る天体の運動理論　ーー多くの観測結果にもっともよく合う軌道の決定ーーという論文（１８０９年）で密度関数を示した。

2　大数の法則

いま，表の出る確率が p, 裏のでる確率が $1-p$ のコインがあったとする。このコインを投げ表が出れば，１点，裏ならば０点として，この点数が表される各回の結果を $X_1, X_2, \cdots, X_n$ とおく。$\Sigma_i X_i$ は１がでた回数で，表の出た比率はデータ平均 $\overline{X}$ となる。大数の法則とは，この表の出た比率は，繰り返し数をどんどん大きくすると，１回投げたときに表れるべき表の確率の値 p に近づくことという命題である。近づくという意味を数学的には，確率収束とか，概収束で正確に書き下される。より一般には,「独立な同一分布にしたがう確率変数列の標本平均（観測値を全体個数で割ったもの，データの平均，算術平均）は，個数を大きくしていくともとの分布の期待値（分布の平均）に確率収束（概収束）する」ということである。この定理は，抽出した標本数を大きくしていけばいくほど，母集団の平均を明らかにすることができるという，統計学の基本定理の一つである。

3　中心極限定理

大数の法則が，標本の平均（観測値の和をデータ総数で割ったもの）と分布の期待値（平均値）との関係を明らかにしているが，中心極限定理とは，同じようにデータ数を大きくしていった状況の近づきかたを示している。つまり，２項分布の極限が正規分布であるということがひとつの例で

ある。繰り返されるコイン投げの列などを例にして，確率変数 $X_1, X_2, \cdots, X_n, \cdots$ は独立ですべて同じ平均 μ, 分散 σ^2 をもつとします。この n 個の観測結果から，標本平均　$\overline{X}_n = \Sigma_{i=1}^n X_i$ を作る。大数の法則から $\overline{X}_n - \mu$ は 0 に確率収束します。これではどのように近づいていくのか分からないので，$\dfrac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}}$ を考える。この分数の分母には，$\overline{X}_n$ の分散の平方，すなわち，標準偏差です。もちろんこの値は n が大きくなれば，ゼロに近づきますが，分子もゼロですから極限の不定形になっています。ラプラスは，この極限分布が正規分布とよばれる分布であることを示しました。数学的に表現すると，任意の実数 a, b に対し，とすれば

$$\lim_{n\to\infty} P\left(a \leq \frac{\overline{X_n} - \mu}{\sqrt{\sigma^2/n}} \leq b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})\, dx$$

が成り立つ。右辺の積分値のなかに表れている関数が標準正規分布の密度関数であり，ガウスが微分方程式で導いたものと同じである。積分の計算がそのままになっている理由は，原始関数が求められない（つまり積分が求められない）から，形を書いているだけである。もし実際に a, b を入れて，この値を求めるには「正規分布表」とよばれる，区間と対応する確率（積分値）を予め数値計算したものを準備しておかねばならない。簡単に言えば，面積の計算が積分で求められないから数値計算したというわけである。この数値表は，統計学のテキストには必ず掲載されているし，今後の推定や検定にはよく用いられるので手元に準備しておく必要がある。

この結果をいいかえると標本平均 $\sqrt{n}(\overline{X}_n - \mu)$ の漸近分布 (個数が大きく場合の分布）は $N(0, \sigma^2)$ である。ここで $N(\mu, \sigma^2)$ とは平均 μ と分散 σ^2 の一般正規分布を表す。

$$\frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \quad -\infty < x < \infty$$

係数は面積が 1 となるようにしたもので，$\dfrac{1}{\sqrt{2\pi}\sigma}$ とか $\dfrac{1}{\sqrt{2\pi\sigma^2}}$ などとも表すので違いに注意する。グラフの概形は，ベル型をした一つ山で，頂点 $x = \mu$ が極大，2 つの変曲点 $x = \mu \pm \sigma$。ここで変曲点とは，上にふくらんだ状態から下にふくらみをもつ状態に変わるところ，極大・極小が 1 階の微分係数をゼロとして求められ，変曲点は凸と凹の境目で，2 階の微分係数をゼロとして求められる。

ガウスの示した中心極限定理のすばらしいことは，もとの分布がコイン投げ（パラメータ p のベルヌーイ分布）であろうがさいころ振り（6 点集合の離散型一様分布）でもカマワナイこと。離散型，連続型のものであっても，極限は連続型の正規分布になる。多少の偏りがあっても，独立な繰り返しとその平均と分散の存在が保証されている限り，どんな分布であっても極限は正規分布になります。非常に強力で有用な定理であることがわかるだろう。

2 項分布の正規近似をおこなうときに，ひとつ注意をすることがある。半整数補正とよばれる近似をよくするためのテクニックである。2 項分布は整数の値をとる離散型で，一方正規分布は実数値をとる連続型であるから，次のような補正をする。

確率変数 X をパラメータ n, p の 2 項分布とするとき，中心極限定理から平均 np, 分散 $np(1-p)$ の正規分布 Y で近似される。また標準正規分布を Z とすると，このばあいの半整数補正とは

$$P(X = k) \fallingdotseq P(k - 0.5 < Y < k + 0.5) = P\left(\frac{(k-0.5) - np}{\sqrt{np(1-p)}} < Z < \frac{(k+0.5) - np}{\sqrt{np(1-p)}}\right),$$
$$P(i \leq X \leq j) \fallingdotseq P(i - 0.5 < Y < j + 0.5) = P\left(\frac{(i-0.5) - np}{\sqrt{np(1-p)}} < Z < \frac{(j+0.5) - np}{\sqrt{np(1-p)}}\right)$$

4　乱数の生成

関数電卓やパソコンの表計算ソフトには，乱数を生成する機能 randomize, rnd(), rand(), ran# などをもつ。シミュレーションや数値計算にはよく用いられる。いわゆる一様乱数とよばれるものであるが，統計・確率の言葉でいえば，一様分布にしたがう確率変数である。厳密に議論するとでたらめとは何かという概念まで深入りしないといけないので，ここでは単に計算機のつくる確率変数を信用しておくことにする。この計算機がつくった確率変数（擬似一様乱数）からさまざまな確率分布のシュミレーションを考えよう。最初に単位区間 $[0,1]$ 上で一様分布する確率変数を $U, U_1, U_2, \cdots$ などとしておく。

１枚のコイン投げ；$X = \begin{cases} 1; & p, \\ 0; & 1-p \end{cases}$ をつくるには，実現値 U が区間 $[0,p]$ の内部か，そうでないかという場合分けをおこなう。

$$\textbf{if } 0 < U < p, \textbf{ let } X = 1, \textbf{ else } X = 0.$$

さいころ振り；$X = i; 1/6, (i = 1,2,3,4,5,6)$ についても上と同様に $[0,1]$ 上で一様分布する確率変数 U から，

$$\begin{aligned} &\textbf{if } 0 \le U < 1/6, \textbf{ let } X = 1, \\ &\textbf{elseif } 1/6 \le U < 2/6, \textbf{ let } X = 2, \\ &\textbf{elseif } 2/6 \le U < 3/6, \textbf{ let } X = 3, \\ &\cdots\cdots \\ &\textbf{elseif } 5/6 \le U < 1\,, \textbf{ let } X = 6. \end{aligned}$$

これらの方法は，逆変換法とよばれるものである。あるいはある数を超えないを対応させる関数 int() をつかうと，**let** $X = int(6 * U) + 1$ でもよいことがわかる。

ｎ枚のコイン投げ；これは２項分布のシミュレーションをすればいいから，１枚のコイン投げを n 回繰り返して，和をとる。これには 100 枚のコイン投げであれば，100 個乱数を作らなければならない。その代わりに，上の逆変換法のように１個の一様乱数から作る方法がある。n, p は予め与えておいてから，

STEP.1 **let** $c = p/(1-p)$, $i = 0$, $pa = (1-p)^n$, $F = pa$.
STEP.2 **if** $U < F$, **let** $X = i$, STOP.
STEP.3 **let** $pa = [c(n-i)/(i+1)] * pa$, $F = F + pa$, $i = i + 1$.
STEP.4 **goto** STEP2.

正規分布にしたがう乱数，正規乱数；中心極限定理の応用として，正規乱数を生成する。これには 12 個の一様乱数から１個の正規乱数がつくれる。

$$\textbf{for } i = 1 \textbf{ to } 12, X = X + U_i \textbf{ next } i, \textbf{ let } X = X - 6$$

つまり，12 個の一様乱数をつけ加えて，これから，６を引くという，極めて簡単なアルゴリズムで，正規乱数が作れる。

5　正規分布の確率計算

正規分布に関する確率を計算するには，正規分布表をもちいる。この表は平均 0，分散 1 の標準正規分布 Z の分布関数を表にまとめたものです。横軸にとり得る値 x（$-\infty$ から ∞）をとり縦軸に

$\Phi(x) = P(Z \leq x)$ を対応させます。密度関数の左右対称性から，正の値のみを書いてあるものが多いですが，これからでもすべての値を計算できます。連続型であることから，$P(Z \leq x) = P(Z < x)$ が成り立つ，離散型とことなり，等号が含まれていてもいなくても同じ値であることに注意します。

いくつかの点を取り上げると，

x	0.000	1.00	1.645	1.960	2.00	2.241	2.326	2.576	2.807	3.00
$\Phi(x)$	.5000	.8413	.9495	.9750	.9772	.9875	.9900	.9950	.9980	.99865

例題

$P(0 < Z < 1.2) = P(0 < Z \leq 1.2) = P(0 \leq Z < 1.2) = P(0 \leq Z \leq 1.2) \fallingdotseq 0.3849$
$P(-1.2 < Z < 0) = P(0 < Z < 1.2) \fallingdotseq 0.3849$
$P(-1.2 < Z < 1) = P(0 < Z < 1.2) + P(0 < Z < 1) \fallingdotseq 0.3849 + 0.3413 = 0.7262$
$P(Z < -1.2) = P(Z < 0) - P(-1.2 < Z < 0) = 0.5 - P(0 < Z < 1.2) \fallingdotseq 0.5 - 0.3849 = 0.1151$

6 練習問題

1 確率変数 X が平均 10, 分散 4 の正規分布にしたがうとき，つぎの値をもとめよ。
(1) $P(X \leq 13)$ (2) $P(X > 11)$ (3) $P(9 < X < 12)$ (4) $P(X < c)$ を満たす c の値

2 ある測定器具の誤差は平均 0, 標準偏差 0.2(分散 $0.2^2 = 0.04$)mm の正規分布にしたがう。この器具で測定誤差が 0.5mm 以上となる確率はいくつか（ヒント；誤差は両側を考えるから，絶対値として一定以上となる確率を求める）

3 ポアソン分布について，$P(X = k+1)$ と $P(X = k)$ の関係式 $(k = 0, 1, 2, \cdots)$ を導き，２項分布と同様なアルゴリズムを作りなさい。

4 コインを 400 回投げるとき，表のでた回数が 180 回以上，210 回以下となる確率はいくつか？半整数補正も加えて，正規分布近似で計算しなさい。

5 中心極限定理の応用として，正規乱数を求める方法の根拠を説明しなさい。（ヒント；和の分布として平均と分散を計算してみなさい）