正規分布の確率密度関数について

林　　雄 一 郎

１　はじめに

二項分布や正規分布は数学 B の確率分布の単元の内容だが、教科書では天下り的な記述にならざるを得ない。例えば、正規分布の確率密度関数には$1/\sqrt{2\pi}\sigma$という係数がなぜついているか、二項分布の確率計算に正規分布表がなぜ使えるかなど疑問がでてくる。これらは授業で厳密に扱えないが、教師はそれを知っておく必要がある。本稿は正規分布の密度関数の性質を詳しく調べ、中心極限定理までを視野にいれ論説することを狙いとする。

２　正規分布の確率密度関数の性質

確率変数Xの値域が実数全体で、Xの確率密度関数$f(x)$が以下の①であるとき、Xの確率分布を平均m、標準偏差σの正規分布$N(m,\sigma^2)$といい、Xは $N(m,\sigma^2)$ に従うという。

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad ・・・①$$

特に、平均０，標準偏差１の正規分布$N(0,1)$を標準正規分布といい、この分布の確率密度関数は①で$m=0,\sigma=1$とおいた②式となる。

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

２－１　②のグラフの形状は？

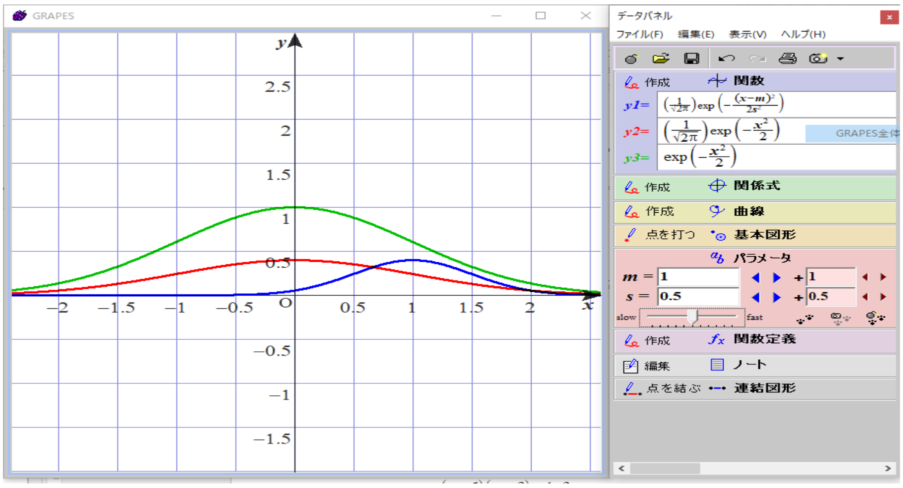
変域は$-\infty<u<+\infty$であり、値域は$g(u)>0$である。$e^{-\frac{u^2}{2}}=\dfrac{1}{e^{\frac{u^2}{2}}}$　であるから

$u\to\pm\infty$　のとき$g(u)\to 0$　となりx軸が漸近線になる。

$g'(u)=-ug(u)=0$　より $u=0$ を境に $g'(u)$ の符号は＋から－に変化するから $u=0$ のとき極大値 $g(0)=\dfrac{1}{\sqrt{2\pi}}\approx 0.399904\cdots$ となる。y 軸が対称軸である。

$g''(u)=-g(u)-ug'(u)=(u^2-1)g(u)=0$　$u=-1,+1$ のとき変曲点を与える。

$u<-1$　では $g''(u)>0$　で下に凸、$-1<u<1$ では $g''(u)<0$ で上に凸、$1<u$ では $g''(u)>0$ で下に凸となる。　以上から類推して、①の正規分布曲線は $x=m$ が対称軸となり、$x=m\pm\sigma$ で変曲点を与える。下図で①は青、②は赤のグラフである。



2－2　関数 $f(x)$、$g(u)$ の性質

$f(x)$ については以下の式が成り立つ。

・全面積　$\int_{-\infty}^{\infty} f(x)dx=1$　・・・③

・平均　$E(X)=\int_{-\infty}^{\infty} xf(x)dx=m$　・・・④

・分散　$\sigma^2(X)=\int_{-\infty}^{\infty}(x-m)^2 f(x)dx=\sigma^2$　・・・⑤

また、$g(u)$ については次の式が成り立つ。

・全面積　$\int_{-\infty}^{\infty} g(u)du = 1$

・平均値　$E(U) = \int_{-\infty}^{\infty} ug(u)du = 0$

・分散　$\sigma^2(U) = \int_{-\infty}^{\infty} u^2 g(u)du = 1$

（証　明）　$f(x)$についての③～⑤を証明する。③の証明は次の変数変換を行って処理する。

$$u = (x-m)/\sqrt{2}\sigma \quad \text{このとき} \quad x = m + \sqrt{2}\sigma u \quad dx = \sqrt{2}\sigma du$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2} \cdot \sqrt{2}\sigma du = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} du = \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-u^2} du \quad \cdot\cdot\cdot ⑥$$

そこで　$\int_{0}^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$　が証明できればいい。この証明は以下、補題を４つ乗り越えねばならない結構な計算量となる。

補題１　$S_n = \int_0^{\pi/2} sin^n xdx$　とおくとき次式が成り立つ。

$$S_n = \frac{(n-1)(n-3)\cdot\cdot 4\cdot 2}{n(n-2)(n-4)\cdot\cdot 3\cdot 1} \cdot \frac{\pi}{2} \qquad (n\text{：偶数})$$

$$= \frac{(n-1)(n-3)\cdot\cdot 4\cdot 2}{n(n-2)\cdot\cdot 3\cdot 1} \qquad (n\text{：奇数})$$

（証　明）　部分積分を行う。

$$S_n = \left[-cos\,x\, sin^{n-1} x\right]_0^{\pi/2} + (n-1)\int_0^{\pi/2} cos^2 x\, sin^{n-2} xdx$$

$$= (n-1)\int_0^{\pi/2} sin^{n-2} xdx - (n-1)\int_0^{\pi/2} sin^n xdx = (n-1)S_{n-2} - (n-1)S_n$$

これから漸化式　$S_n = \frac{n-1}{n} S_{n-2}$　（$n \geq 2$）　を得るから成り立つ。

$S_0 = \int_0^{\pi/2} dx = \frac{\pi}{2}$　　$S_1 = \int_0^{\pi/2} sin\,xdx = \left[-cos\,x\right]_0^{\pi/2} = 1$　となる。

例えば　$n=6$　では　$S_6=\dfrac{5\cdot 3\cdot 1}{6\cdot 4\cdot 2}\cdot\dfrac{\pi}{2}=\dfrac{15\pi}{96}$

$n=7$　では　$S_7=\dfrac{6\cdot 4\cdot 2}{7\cdot 5\cdot 3\cdot 1}=\dfrac{48}{105}$　となる。

n が偶数のとき $\dfrac{\pi}{2}$ がかかる。

補題２　$\int_0^1\left(1-x^2\right)^n dx=S_{2n+1}$　・・⑦　　$\int_0^\infty \dfrac{1}{\left(1+x^2\right)^n}dx=S_{2n-2}$　・・⑧

が成り立つ。

（証　明）

$x=\cos t$ とおくと⑦の左辺は、

$$\int_{\pi/2}^0\left(1-\cos^2 t\right)^n\left(-\sin t\right)dt=\int_0^{\pi/2}\sin^{2n+1}dt=S_{2n+1}$$

次に、$x=\tan t$　とおくと　$1+\tan^2 t=1/\cos^2 t$　だから⑧の左辺は

$$\int_0^{\pi/2}\cos^{2n}\sec^2 tdt=\int_0^{\pi/2}\cos^{2n-2}tdt \quad ・・⑨$$

$t=\dfrac{\pi}{2}-s$　とおけば③は　$\int_{\pi/2}^0\cos^{2n-2}\left(\dfrac{\pi}{2}-s\right)\left(-ds\right)=\int_0^{\pi/2}\sin^{2n-2}sds=S_{2n-2}$

補題３

$$\lim_{n\to\infty}\frac{S_{2n}}{S_{2n+1}}=1$$　が成り立つ。

（証　明）

$0<x<\pi/2$　ならば　$0<\sin^{n+1}x<\sin^n x$　だから $\left\{S_n\right\}$ は単調減少数列である。

$$1<\frac{S_{2n}}{S_{2n+1}}<\frac{S_{2n-1}}{S_{2n+1}}=\frac{(2n-2)(2n-4)\cdot\cdot 4\cdot 2}{(2n-1)(2n-3)\cdot\cdot 3\cdot 1}\div\frac{2n(2n-2)(2n-4)\cdot\cdot 4\cdot 2}{(2n+1)(2n-1)\cdot\cdot 3\cdot 1}=\frac{2n+1}{2n}\to 1$$

$$\frac{S_{2n}}{S_{2n+1}}=\frac{(2n-1)(2n-3)\cdot\cdot 3\cdot 1}{2n(2n-2)\cdot\cdot 4\cdot 2}\cdot\frac{\pi}{2}\div\frac{2n(2n-2)\cdot\cdot 4\cdot 2}{(2n+1)(2n-1)\cdot\cdot 3\cdot 1}=\frac{\pi}{2}\cdot\prod_{k=1}^{n}\frac{(2k+1)\cdot(2k-1)}{2k\cdot 2k}\to 1$$

この式から　$\displaystyle\prod_{k=1}^{\infty}\frac{2k\cdot 2k}{(2k+1)(2k-1)}=\frac{\pi}{2}$　あるいは

$$\prod_{k=1}^{\infty}\frac{(2k+1)(2k-1)}{(2k)^2}=\prod_{k=1}^{\infty}\left(1-\frac{1}{(2k)^2}\right)=\frac{2}{\pi}\qquad \text{（Wallis の公式）}$$

$\displaystyle\frac{3}{4}\cdot\frac{15}{16}\cdot\frac{35}{36}\cdot\frac{63}{64}\cdot\frac{99}{100}\cdots\to\frac{2}{\pi}$　　不思議な式である。どうやって見つけたのであろう？！

補題４　$\displaystyle\lim_{n\to\infty}\sqrt{n}S_{2n+1}=\frac{\sqrt{\pi}}{2}$

（証　明）

$$S_{2n}S_{2n+1}=\frac{(2n-1)(2n-3)\cdots 3\cdot 1}{2n(2n-2)\cdots 4\cdot 2}\cdot\frac{\pi}{2}\cdot\frac{2n(2n-2)\cdots 4\cdot 2}{(2n+1)(2n-1)\cdots 3\cdot 1}=\frac{\pi}{2(2n+1)}$$

$\displaystyle\sqrt{S_{2n}S_{2n+1}}=\sqrt{\frac{\pi}{2(2n+1)}}=S_{2n+1}\sqrt{\frac{S_{2n}}{S_{2n+1}}}$　この式から $\displaystyle S_{2n+1}=\sqrt{\frac{S_{2n+1}}{S_{2n}}}\cdot\sqrt{\frac{\pi}{2(2n+1)}}$

したがって　$\displaystyle\sqrt{n}S_{2n+1}=\sqrt{\frac{n\pi}{2(2n+1)}}\cdot\sqrt{\frac{S_{2n+1}}{S_{2n}}}\to\frac{\sqrt{\pi}}{2}$　　・・・⑩

ところで　$x\neq 0$　ならば　$e^x>x+1$　だから　$e^{-x^2}>1-x^2$　$e^{x^2}>1+x^2$

よって　$\displaystyle e^{-x^2}<\frac{1}{1+x^2}$　　ゆえに　$\displaystyle(1-x^2)^n<e^{nx^2}<\frac{1}{(1+x^2)^n}$

$$S_{2n+1}=\int_0^1(1-x^2)^n dx<\int_0^1 e^{-nx^2}dx<\int_0^{\infty}e^{-nx^2}dx<\int_0^{\infty}\frac{dx}{(1+x^2)^n}=S_{2n-2}$$

他方　$\displaystyle x=\frac{t}{\sqrt{n}}$　とおくと　$\displaystyle\int_0^{\infty}e^{-nx^2}dx=\int_0^{\infty}e^{-t^2}\cdot\frac{dt}{\sqrt{n}}=\frac{1}{\sqrt{n}}\int_0^{\infty}e^{-x^2}dx$

したがって　$\displaystyle\sqrt{n}S_{2n+1}<\int_0^{\infty}e^{-x^2}dx<\sqrt{n}S_{2n-2}$　　④から

$n \to \infty$　のとき　$\displaystyle\int_0^{\infty} e^{-x^2} dx \to \frac{\sqrt{\pi}}{2}$　となる。

（別　解）　二重積分を使うと簡単に求められるがこれは 2 変数の微積分の知識が必要となる。ちなみに計算は

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx = \int_{-\infty}^{\infty} e^{-y^2} dy \quad \text{とおく。} \quad I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2} dy\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\left(x^2+y^2\right)} dxdy$$

・・⑪

$x = r\cos\theta, y = r\sin\theta$　と変数変換する。

$$⑪ = \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2} rdr = 2\pi\left[-\frac{1}{2}e^{-r^2}\right]_0^{\infty} = \pi \qquad \text{よって} \quad I = \sqrt{\pi}$$

次に、平均　$E\left(X\right) = \int_{-\infty}^{\infty} xf\left(x\right)dx = m$　の証明である。

$$\int_{-\infty}^{\infty} xf\left(x\right)dx = \int_{-\infty}^{\infty}\left(x-m\right)f\left(x\right)dx + m\int_{-\infty}^{\infty} f\left(x\right)dx = \left[-\frac{\sigma}{\sqrt{2\pi}}e^{-\frac{\left(x-m\right)^2}{2\sigma^2}}\right]_{-\infty}^{\infty} + m = m$$

最後に、分散の計算は以下の通りである。

$$V\left(X\right) = \int_{-\infty}^{\infty}\left(x-m\right)^2 f\left(x\right)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\left(x-m\right)\cdot\left(x-m\right)e^{-\frac{\left(x-m\right)^2}{2\sigma^2}}dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\left[\left(x-m\right)\cdot\left(-\sigma^2\right)\left(e^{-\frac{\left(x-m\right)^2}{2\sigma^2}}\right)\right]_{-\infty}^{\infty} + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{\left(x-m\right)^2}{2\sigma^2}}dx = \sigma^2\int_{-\infty}^{\infty} f\left(x\right)dx = \sigma^2$$

ここで $\displaystyle\lim_{x\to\pm\infty}\frac{x-m}{e^{\frac{\left(x-m\right)^2}{2\sigma^2}}} = \lim_{y\to\pm\infty}\frac{1}{\frac{y}{\sigma^2}e^{\frac{y^2}{2\sigma^2}}} = 0$

変数変換 $y = x - m$ を施して分母分子を微分し極限をとってロピタルの定理を用いている。

２－３　$f\left(x\right)$　と $g\left(u\right)$　との関係は次式である。

$$\int_{m+a\sigma}^{m+b\sigma} f\left(x\right)dx = \int_a^b g\left(u\right)du \qquad ・・・⑫$$

（証　明）

$u = (x-m)/\sigma$　という変数変換を行う。$x = m + \sigma u$　$dx = \sigma du$

x の値が $m+\sigma a$ 、$m+\sigma b$ のとき u の値はそれぞれ a,b となる。

$$⑫ = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{u^2}{2}} \cdot \sigma du = \int_a^b g(u) du$$

この式の意味は確率変数 X の確率分布は $U = (X-m)/\sigma$ なる確率変数 U の確率分布となることをいっている。確率変数 X が正規分布 $N(m,\sigma^2)$ に従うとき、確率変数 U は標準確率分布 $N(0,1)$ に従うことを意味する。

例えば、$P(m+\sigma a \le X \le m+\sigma b) = P(a \le U \le b)$ である。これが U の確率分布（正規分布表）を用いて計算できる根拠となる。

例　X が正規分布 $N(3,25)$ に従うとき、$X \le 7$ となる確率を計算せよ。

$m = 5, \sigma = 25$ だから $U = (X-m)\sigma = (7-3)/5 = 0.8$

$P(X \le 7) = P(U \le 0.8) = 0.5 + 0.2881 = 0.7881$

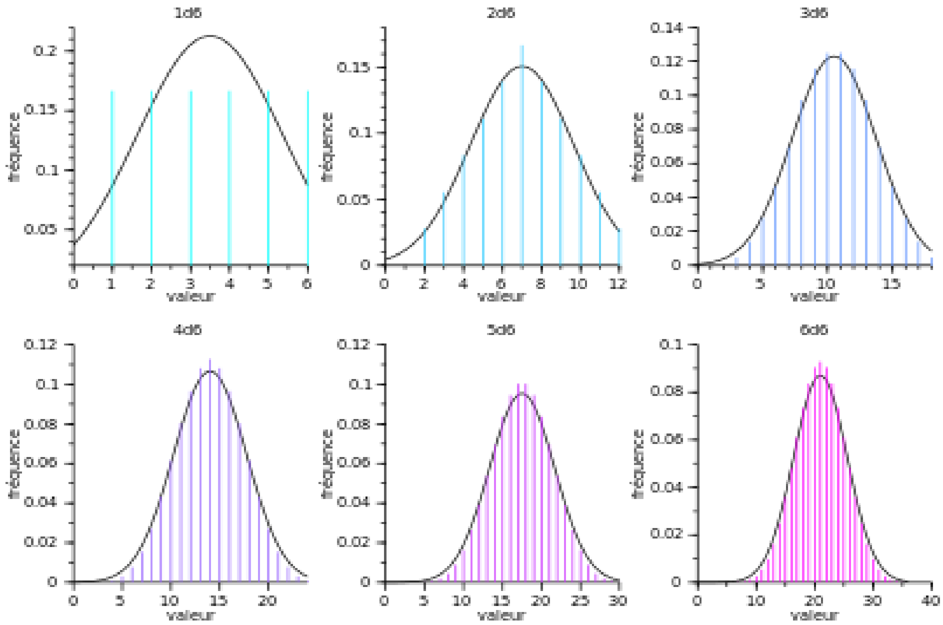
２－４　二項分布と正規分布の関係

確率変数 X が二項分布 $B(n,p)$ に従うとき、n の値が大きければ $U = (X-np)/\sqrt{npq}$（$p+q=1$）はほぼ標準正規分布 $N(0,1)$ に従う。n を大きくしたとき $B(n,p)$ は $N(np,npq)$ に近づくという定理（De-Moivre,Laplace の定理）が成り立つ。すなわち、

$$\lim_{n\to\infty} P(U \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad ・・・⑬$$

これから　$P(\alpha \le U \le \beta) \to \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$　が出てくる。

教科書にはこの定理の証明は無く使い方の解説のみがある。この定理は二項分布の標準正規分布近似についてのもので正規分布表を用いて計算できる根拠を与える。



このグラフはサイコロを n 回振ったときの目の和の分布であり、n を大きくすると正規分布に近づくことを示すものである。

なお、この定理を一般化した定理が以下の中心極限定理である。

2－5　中心極限定理とは

確率変数 $X_1, X_2, \cdots, X_n$ がすべて独立で、同じ分布（二項分布とは限らない！）に従うとする。すなわち、$E\left(X_i\right) = m, V\left(X_i\right) = \sigma^2$ とする。

このとき、標本平均 $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ の平均、分散は次式で与えられる。

$$E\left(\overline{X}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\cdot nm = m = E\left(X_i\right)$$

$$V\left(\overline{X}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n} = \frac{1}{n}V\left(X_i\right)$$

また、基準化された確率変数 $\dfrac{\overline{X} - m}{\sqrt{V\left(\overline{X}\right)}} = \dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$ の平均、分散は

$$E\left(\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}\right)=\frac{\sqrt{n}}{\sigma}E\left(\overline{X}\right)-\frac{m\sqrt{n}}{\sigma}=0$$

$$V\left(\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}\right)=\frac{n}{\sigma^2}V\left(\overline{X}\right)=\frac{n}{\sigma^2}\cdot\frac{\sigma^2}{n}=1$$

となり、基準化でその期待値、分散がそれぞれ０，１となる。

中心極限定理

n が十分大きいとき $\dfrac{\overline{X}-m}{\sqrt{V\left(\overline{X}\right)}}=\dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ は標準正規分布 $N(0,1)$ に従う。

そして次式がなりたつ。

$$\lim_{n\to\infty}P\left(\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}\leq x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du \qquad \cdots ⑭$$

この証明は一般の確率変数を扱うので難解であり統計学の専門書を参照されたい。
そこで、二項分布に従う確率変数に成り立つ De-Moivre,Laplace の定理で我慢するのだがこの証明は中心極限定理を用いれば以下のように容易に導かれる。

（証　明）

確率変数 X は n 回行って事象が起こる回数で二項分布に従うから

$$E\left(X_i\right)=1\cdot p+0\cdot\left(1-p\right)=p \quad V\left(X_i\right)=E\left(X_i^2\right)-\left(X\left(X_i\right)\right)^2=1^2p-p^2=p\left(1-p\right)$$

となる。また、$X=X_1+X_2+\cdots+X_n$ とすれば

各 X_i の起こる確率は p だから $E\left(X\right)=\sum_{i=1}^{n}E\left(X_i\right)=np$

各 X_i は独立だから $\sigma^2=V\left(X\right)=\sum_{i=1}^{n}V\left(X_i\right)=np\left(1-p\right)$

だから $\overline{X}=X/n \quad p=E\left(\overline{X}\right)=p,V\left(\overline{X}\right)=\sigma^2/n=p\left(1-p\right)$

$$\frac{\overline{X}-p}{\sigma/\sqrt{n}}=\frac{X/n-p}{\sqrt{p\left(1-p\right)}}\cdot\sqrt{n}=\frac{X-np}{\sqrt{np\left(1-p\right)}}$$ ⑭に代入すれば証明される。

２－６　De-Moivre,Laplace の定理の証明

　二項分布の式には階乗が出てくる。これを処理するには Stirling の公式

$n\to\infty$ のとき $n!:\ \sqrt{2\pi}n^{n+\frac{1}{n}}e^{-n}$　が必要（証明は後回し）となる。

(1)　二項分布 $B\left(n,p\right)$ に従う確率変数 X が k となる確率 $P\left(X=k\right)={}_nC_k\,p^kq^{n-k}$ は

$p+q=1$　$\left(k-np\right)/\sqrt{npq}=\lambda<\infty$ が成り立つとき、$n\to\infty$ ならば $\dfrac{1}{\sqrt{2\pi npq}}e^{-\frac{\lambda^2}{2}}$ で近似できる。

（証　明）　$\lambda<\infty$□のとき　$k=np+\lambda\sqrt{npq}=np\left(1+\lambda\sqrt{\dfrac{q}{np}}\right)\to\infty$

同様に　$n-k=n\left(p+q\right)-np-\lambda\sqrt{npq}=nq\left(1-\lambda\sqrt{\dfrac{p}{nq}}\right)\to\infty$

　ここで Stirling の公式を用いる。

$$n!\sim\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n},k!\sim\sqrt{2\pi}k^{k+\frac{1}{2}}e^{-k},\left(n-k\right)!\sim\sqrt{2\pi}\left(n-k\right)^{n-k+\frac{1}{2}}e^{-\left(n-k\right)}$$

$${}_nC_k\,p^kq^{n-k}\sim\frac{1}{\sqrt{2\pi}}\cdot\frac{n^{n+\frac{1}{2}}}{k^{k+\frac{1}{2}}\left(n-k\right)^{n-k+\frac{1}{2}}}p^kq^{n-k}=\frac{1}{\sqrt{2\pi}}AB$$ とすれば

$$A=\sqrt{\frac{n}{k\left(n-k\right)}}=\sqrt{\frac{n}{\left(np+\lambda\sqrt{npq}\right)\left(nq-\lambda\sqrt{npq}\right)}}=\sqrt{\frac{n}{n^2pq\left(1+\lambda\sqrt{\frac{q}{np}}\right)\left(1-\lambda\sqrt{\frac{p}{nq}}\right)}}$$

$$=\frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{\left(1+\lambda\sqrt{\frac{q}{np}}\right)\left(1-\lambda\sqrt{\frac{p}{nq}}\right)}}\approx\frac{1}{\sqrt{npq}}$$

$$B=\left(\frac{np}{k}\right)^{k}\left(\frac{nq}{n-k}\right)^{n-k}=\left(\frac{np}{np+\lambda\sqrt{npq}}\right)^{k}\left(\frac{nq}{nq-\lambda\sqrt{npq}}\right)^{n-k}$$

$$=\left(\frac{1}{1+\sqrt{\frac{q}{np}}}\right)^{k}\left(\frac{1}{1+\sqrt{\frac{p}{nq}}}\right)^{n-k}$$

$$\log B=-k\log\left(1+\lambda\sqrt{\frac{q}{np}}\right)-\left(n-k\right)\log\left(1-\lambda\sqrt{\frac{p}{nq}}\right)$$

$$=-\left(np+\lambda\sqrt{npq}\right)\log\left(1+\lambda\sqrt{\frac{q}{np}}\right)-\left(nq-\lambda\sqrt{npq}\right)\log\left(1-\lambda\sqrt{\frac{p}{nq}}\right)=*$$

ここで　$\log\left(x+1\right)\approx x-\frac{x^2}{2}$　（Taylor 展開）を用いれば

$$*=-\left(np+\lambda\sqrt{npq}\right)\left(\lambda\sqrt{\frac{q}{np}}-\frac{\lambda^2 q}{2np}\right)-\left(nq-\lambda\sqrt{npq}\right)\left(-\lambda\sqrt{\frac{p}{nq}}-\frac{\lambda^2 p}{2nq}\right)$$

$$=-\lambda\sqrt{npq}+\frac{\lambda^2 q}{2}-\lambda^2 q+\frac{\lambda^3 q}{2}\sqrt{\frac{q}{np}}+\lambda\sqrt{npq}+\frac{\lambda^2 p}{2}-\lambda^2 p-\frac{\lambda^3 p}{2}\sqrt{\frac{p}{nq}}$$

$$\approx\frac{\lambda^2}{2}\left(p+q\right)-\lambda^2\left(p+q\right)=-\frac{\lambda^2}{2}$$

よって　$B\approx e^{-\frac{\lambda^2}{2}}$　となり　${}_nC_k p^k q^{n-k}\approx\frac{1}{\sqrt{2\pi npq}}e^{-\frac{\lambda^2}{2}}$

例　1 回の試行で事象 E の起こる確率は 0.1 である。この独立試行を 100 回行うとき事象 E が 15 回起こる確率を求めよ。

事象 E が起こる回数Xは 2 項分布 B(100,0.1)に従う。$E\left(X\right)=100\times0.1=10$

$\sigma=\sqrt{100\times0.1\times0.9}=3$　よって確率は　${}_{100}C_{15}\left(0.1\right)^{15}\left(0.9\right)^{95}$　となる。

この値は$0.032682\cdot\cdot$（関数電卓 TI-Nｓｐｉｒｅを使用）

この値を標準正規分布$N\left(10,0.1^2\right)$で求めるには、少し幅を持たせて$14.5<X<15.5$

となる確率を求める。

$U = \dfrac{X - 10}{3}$ とおくと、正規分布表を使って

$$P\left(14.5 < X < 15.5\right) = P\left(1.5 < X < 1.83\right) = \int_{1.5}^{1.83} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$= \int_0^{1.83} - \int_0^{1.5} = 0.4664 - 0.4332 \approx 0.033$$

ほぼ妥当な近似値となることが分かる。

（２）次に、区間$\left[\alpha,\beta\right]$を長さ$\dfrac{1}{\sqrt{npq}}$の等間隔で$s$等分する。

$$\frac{s}{\sqrt{npq}} = \beta - \alpha \quad s = \left(\beta - \alpha\right)\sqrt{npq} \to \infty \quad （n \to \infty）$$

分点を $\alpha = x_0 < x_1 < \cdots < x_{s-1} < x_{s-1} < x_s = \beta$ とする。 $y = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ のグラフと

$x = x_i, x = x_{i+1}$、x軸で囲まれた矩形図形全体の面積の和をS_nとおくと

$$S_n = \sum_{i=1}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \cdot \frac{1}{\sqrt{npq}} \to \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad （n \to \infty）$$

したがって、

$$P\left(a \le X \le b\right) = \sum_{k=a}^{b} {}_nC_k p^k q^{n-k} = P\left(\alpha \le \frac{X - np}{\sqrt{npq}} \le \beta\right) = P\left(\alpha \le U \le B\right)$$

$$\approx \sum_{i=1}^{s} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_i^2}{2}} \to \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

２－７　Stirling の公式の証明

下図は$y = \log x$のグラフである。x軸上の点$1, 1/2, 2, 5/2, \cdots i, i + \dfrac{1}{2}, i + 1, \cdots, n$・・・

での垂線とグラフ、点$\left(i, \log i\right)$でx軸に平行な線分で囲まれる部分の面積を考える。

例えば、$y = \log x$ と $x = 3, x$ 軸で囲まれる図形の面積は

$$\alpha_1 + \left(\log 2 - \beta_1\right) + \alpha_2 + \left(\frac{1}{2}\log 3 - \beta_2\right) = \log 2 + \frac{1}{2}\log 3 + \delta_3 = \int_1^3 \log x dx$$

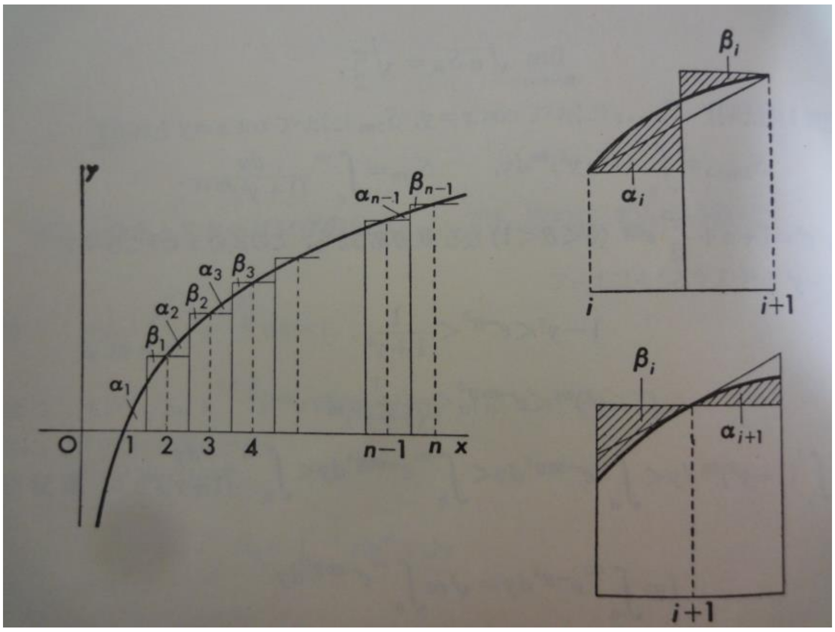
ただし、$\delta_3 = \alpha_1 - \beta_1 + \alpha_2 - \beta_2$ とおく。一般に

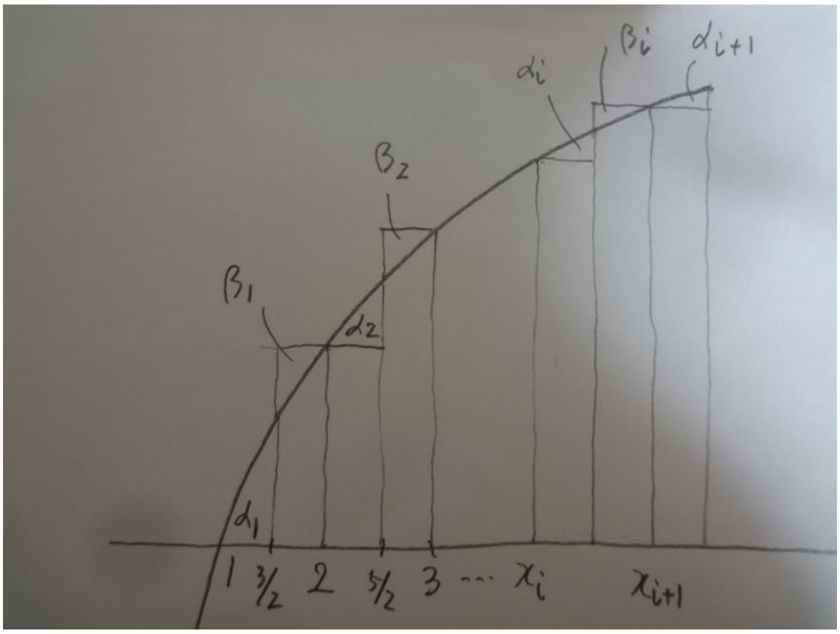
$$\log 2+\log 3+\cdots+\log(n-1)+\frac{1}{2}\log n+\delta_n=\int_1^n \log x\,dx=n\log n-n+1$$

このとき$\delta_n=\alpha_1-\beta_1+\alpha_2-\beta_2+\cdots+\alpha_{n-1}-\beta_{n-1}$　とおく。

$$\log n!=\left(n+\frac{1}{2}\right)\log n-n+1-\delta_n=\log n^{n+\frac{1}{2}}e^{-n}e^{1-\delta_n} \qquad n!=n^{n+\frac{1}{2}}e^{-n}e^{1-\delta_n} \quad \cdots ⑮$$

ここで$\alpha_i>\beta_i>\alpha_{i+1}$　となるから$\alpha_1>\beta_1>\alpha_2>\beta_2>\cdots>\alpha_n>\beta_n$





$\alpha_i > \beta_i > \alpha_{i+1} > \cdot\cdot \to 0$　数列$\{\delta_n\}$は交項級数で$\{\delta_{2n+1}\}$は増加数列、$\{\delta_{2n}\}$は減少数列

$\delta_{2n} - \delta_{2n-1} = \alpha_{n+1} \to 0$　区間$[s_{2n-1}, s_{2n}]$の長さは０に収束するから$\{\delta_n\}$はある極限値δに収束する。$\delta_n \to \delta$　$(n \to \infty)$　補題４から

$$\sqrt{\pi} = \lim_{n\to\infty} 2\sqrt{n} S_{2n-1} = \lim_{n\to\infty} 2\sqrt{n} \cdot \frac{(2n-2)\cdot\cdot 4 \cdot 2}{(2n-1)(2n-3)\cdot\cdot 3 \cdot 1}$$

$$= \lim_{n\to\infty} \frac{n!\,2^n}{\sqrt{n}} \cdot \frac{2n(2n-2)(2n-4)\cdot\cdot 4 \cdot 2}{2n(2n-1)!} = \lim_{n\to\infty} \frac{(n!)^2\, 2^{2n}}{(2n)!\sqrt{n}}$$

$$= \lim_{n\to\infty} \frac{e^{2(1-\delta_n)} n^{2n+1} e^{-2n}\, 2^{2n}}{e^{1-\delta_{2n}}\, 2^{2n+\frac{1}{2}} n^{2n+\frac{1}{2}} e^{-2n} \sqrt{n}} = \lim_{n\to\infty} \frac{e^{1-\delta_n}}{\sqrt{2}}$$

よって　$e^{1-\delta} \approx \sqrt{2\pi}$　　⑮に代入すると　　$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$　を得る。

例えば　$n = 10$　のとき$10! = 3628800$

他方　$10^{10+\frac{1}{2}} = 31622776601$　$e^{-10} \approx 0.0000454$　$\sqrt{2\pi} \approx 2.5059928$

したがって、$\sqrt{2\pi}\, 10^{10+\frac{1}{2}} e^{-10} = 3598695.6189$　誤差は　30105程度である。
$n = 10$くらいでは誤差が大きいことが分かる。

３　あとがき

正規分布の密度関数のいろいろな性質を調べる中で 2 項分布とよく似た確率密度関数の関連性を考察しめんどうな計算を経て De-Moibve,Laplace の定理とその証明まで達した。

さらに、その一般化である中心極限定理まで言及した。その証明はハイレベルな統計解析の知識を必要とするのでお預けとしたが、この論考を理解すればいくらか関数e^{-x^2}に親近感がわいたような気になると思う。