数式ギャラリー (1) 正規分布曲線

SHIMURA Masato

2016 年 12 月 15 日

目次

　一般に数学は到達主義であり、そこで説明されている数式を天下り的に受け入れ、スクリプトを作成することに傾倒していたが、時間も無限に用いることができるので数学史を踏まえ数式の源流をたどりながら、数式を簡潔に記述できる J 言語の特色を生かしスクリプトで数式を縦横に動かすことを試みてみる。

1　正規分布

正規分布の確率密度関数（CDF)

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

1.1　正規分布ができるまで

19 世紀まで学術書はラテン語や時にはフランス語で書かれることが多く、数学史に関する部分をネットに溢れる論文や資料を見て調べてもも、引用が多く、どれがオリジナルか判然としないことが多い。

正規分布曲線はド・モアブル、ラプラス、ガウスの 3 人により今日の形にまとめられた。

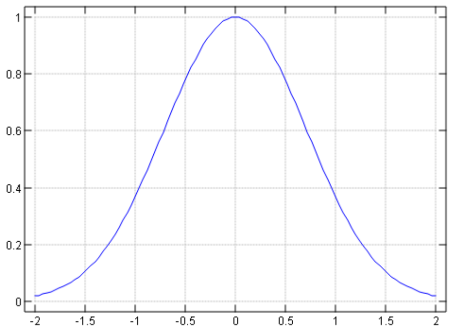
後年カール・ピアソンがガウスとラプラスの分布曲線の優先順位を避けるために、正規分布と呼んだようだ。

ド・モアブル (1667-1754)　.

ド・モアブル (1667-1754) はフランス・シャンパンニュー地方の生まれ。カルバン派で、ルイ 14 世によりナントの勅令が廃棄されたとき多くのユグノーはフランスからオランダなどへ移り住み、フランスの停滞を招いた。ド・モアブルも 18 歳でイギリスに亡命した。統計家でギャンブラーのコンサルタントも。

ド・モアブルは 1733 に次の式を発見した。シンプルで美しい数式から、ベルカーブが現れる。

$$f(x) = e^{-x^2}$$




```
plot tmp; 1x1 ^ -@^&2 tmp=. steps _2 2 100
plot _3 3; '1x1 ^ -@^&2'  NB. Science Plot

pd 'save png c:/temp/d_moibre.png'
```

ラプラス (1749-1827)　.

フランスはノルマンディ生れでモーツアルトより 3 歳年下。王政下で砲兵隊の面接官のとき 16 歳のナポレオンを知る。ナポレオンのエジプト遠征に同行したり大臣にしてもらったが彼の没落に付き合うお人よしではなかった。

ド・モアブル=ラプラスの定理　ド・モアブルは今日「スターリングの公式」と呼ばれるべき乗の近似計算法も開発して、二項分布のモデル数が大きくなると正規分布で近似できることを証明した

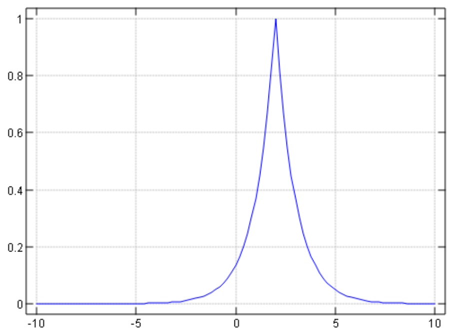
更にラプラスが精緻化した。*Central limit theorem* も見出している

ラプラスの誤差解析　ラプラスは正規分布を用いて誤差解析を行った。（*1778*）

ラプラス分布　データ分布が正規分布より鋭い分布。金融工学や生物学で使われる。

$$\frac{1}{2\phi}e^{-\frac{|x-\mu|}{\phi}}$$

平均　μ　：分散　$2\,\phi^2$　：標準偏差 $\sqrt{2}\phi$

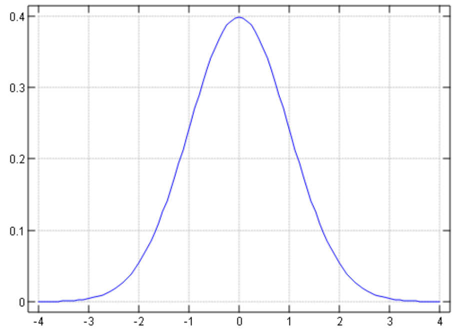



```
mean=: +/ % #

laplace=: 4 :  '(% 1r2 * x) * 1x1 ^  - |(y - mean y) - x'
plot_laplace=: 4 : ' plot y; x laplace y'

2  plot_laplace steps _10 10 100
pd 'save png c:/temp/laplace0.png'
```

ラプラスは *1783* 年に正規分布にほぼ達していたようだ。

ガウス　誤差論や最小二乗法の検討過程で次の式によって一つの分布が定まることを発見している *(1809)*

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$




```
2 plot_gauss steps _4 4 100
pd 'save png c:/temp/gauss_normal0.png'
```

1.2　正規分布の導出

2 項分布は計算が大変で、データの規模が大きくなるとスターリングの公式を用いたりする。また、正規分布で近似できる。

正規分布の導出は難解である。以下の概略は常盤台/統計学の WIKI によった。

テーラー展開　ド・モアブルの式をを $x = 0$ の周りでテーラー展開する

$$f(x) \sim e^{-x^2} = 1 - \frac{1}{1!}x^2 + \frac{1}{2!}x^4 + \cdots + (-1)n\frac{1}{n!}x^{2n} + \cdots$$

規格化　$\pm\infty$　の範囲で積分値が 1 となるように規格化する

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

σ　左右の広がりを表す標準偏差 σ を加える

$$x \to \frac{x}{2\sqrt{\pi}\sigma}$$

整理　.

$$f'(x) = \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-x^2}{2\sigma^2}\right)}$$

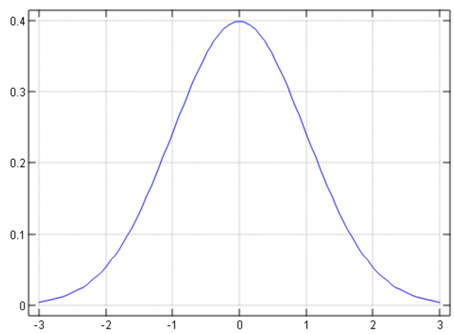
1.3　正規分布のスクリプト

鈴木義一郎「J 言語による統計解析」のスクリプト。

```
ndens=: 3 : '(^--:*:y)%%:o.2'
nden=: 4 : '(ndens(y-{.x)%s)%s=. %:{:x'
```

片側形の *ndens* は標準正規分布で、両側形 *nden* は左引数に平均と分散を与えたもの

```
   ndens 0 0.5 1 2
0.398942 0.352065 0.241971 0.053991
   1 4 nden 1 2 3 4
0.199471 0.176033 0.120985 0.0647588
```




```
plot _3 3 ;'ndens'
```

2 誤差関数 erf

$$erf(a) = \frac{2}{\sqrt{\pi}} \int_0^a e^{-t^2} dt$$

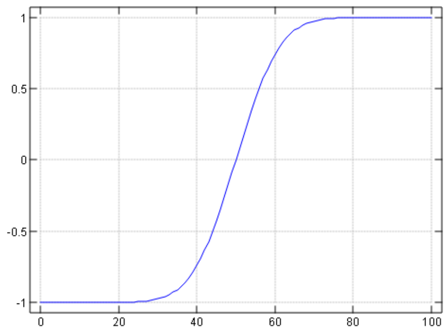
これもガウスによる。正規分布の妹のような分布である。

正規分布に従う測定値の標準偏差が σ, 期待値が *0* の場合に、一つの測定値が $-a \leftrightarrow a$ の間になる確率は次のようになる

$$erf\left(\frac{a}{\sigma\sqrt{2}}\right)$$

```
gamma=:!&<:!
erf   =: (1 H. 1.5)@*: * 2p_0.5&* % ^@:*:
n01cdf=: -: @ >: @ erf @ %&(%:2)
```

```
plot erf steps _4 4 100
```



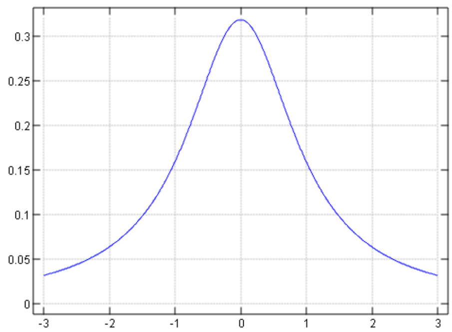
3 コーシー分布

- コーシー（*1789-1857*）による。コーシーはガウスより *12* 歳若く、フランスの動乱期を生きた。革命より避難していたパリ郊外の小さな村で、ラプラスとコーシーの父が知り合いで、ラプラスに擁護された。イエズス会系であり、オペラのロッシーニと同世代
- X, Y が独立で、共に正規分布 $N(0,1)$ に従うとき、$\dfrac{X}{Y}$ の確率密度関数はコーシー分布となる
- 共鳴を表す微分方程式の解になることから物理学で用いられる。
- 正規分布より腰が広い。

- 期待値なく、中心極限定理が当てはまらないことでも知られる。
- 標本数を増やしても平均が *0* に近づく以上に外れ値を取る確率が高く、金融工学でも用いられるようだ

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

```
require 'plot numeric trig png'
cauchy0=: 3 : '% 1p1 *  >: ^&2 y'

 plot tmp;  cauchy0 tmp=. steps _3 3 1000
 pd 'save png c:/temp/cauchy0.png'
```



付録 A　正規分布の導出

次は一石賢「道具としての統計解析」によった。
[*1]

- 次のように置いて、積分する。(二重積分）

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx$$

- $I = \int_{-\infty}^{\infty} e^{-y^2} dy$ としても同じだから、これを両辺に掛け合わせる

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy$$

- 極座標に変換して積分する (極座標での二重積分）

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy$$
$$= \int_{0}^{2\pi} \int_{0}^{\infty} e^{-r^2} rdrd\theta$$
$$= \int_{0}^{2\pi} d\theta \int_{0}^{\infty} re^{-r^2} dr$$
$$= [\theta]_0^{2\pi} \left[-\frac{1}{2} e^{-r^2} \right]_0^{\infty}$$
$$= 2\pi \left\{ 0 - \left(\frac{1}{2} \right) \right\}$$
$$= \pi$$

$$I = \sqrt{\pi}$$

[*1] 数式マニアが優しいふりをして表した怪書である

$\int_{-\infty}^{\infty} e^{-x^2}$ は $\sqrt{\pi}$ となる

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

単位円盤の面積は、半径 *1* の円では π となる。

- ±∞ の範囲で積分値が *1* となるように規格化できた
- 確率変数の標準偏差が σ となるように調整する

$$x \to \frac{x}{\sqrt{2}\sigma}$$

- 平均を *0* ではなく任意の μ とすると与式になる

References

一石賢「道具としての統計解析」日本実業出版社　*2004*

鈴木義一郎「*J* 言語による統計分析」森北出版　*1996*