# No.12 正規分布とその性質

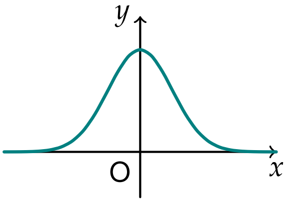
NAKADA Masayuki

Kobe University Secondary School

October 22, 2021

# 正規分布（再掲）

正規分布はさまざまな現象に自然に現れる分布であるのみならず，十分大きいデータサイズの標本の標本平均の分布が正規分布に近づいていく（**中心極限定理**）など，統計学における親玉的分布である．別名ガウス分布．




確率密度関数 $\boxed{f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty)}$

平均と分散 $\boxed{E(X)=\mu,\ \ V(X)=\sigma^2}$

ここで，$e = 2.7182\ldots$ はネイピア数と呼ばれ，極限 $e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$ で与えられる無理数である．特に，$\mu = 0$，$\sigma = 1$ である正規分布 $\mathrm{N}(0, 1^2)$ を**標準正規分布**と呼ぶ．他の正規分布は，確率変数の標準化 $Z = \dfrac{X-\mu}{\sigma}$ によって，標準正規分布に帰着させて計算する．

正規分布に従うとされる例は，多岐に亘る．

- 体長，体重など生物個体に関する測定量
- 雨粒や砂粒の大きさなど，自然生成物に関する測定量
- 学業成績など（状況によっては怪しいこともあるけど）
- 工業生産物等の誤差
- 測定誤差

$\rightsquigarrow$ 正規分布の性質を利用して分析可能！

# 変数の標準化

## 正規分布の性質

正規分布は，平均 $\mu$ と分散 $\sigma^2$（または標準偏差 $\sigma$）によってその分布が決定する．

そこで，平均 $\mu$，標準偏差 $\sigma$ である正規分布 $\mathrm{N}(\mu, \sigma^2)$ に従う確率変数 $X$ の分布の様子を調べるために，**確率変数の標準化**を行い，標準正規分布上で分析し，結果を元の分布に翻訳することを考える．

雑多...

$X \sim \mathrm{N}(\mu, \sigma^2)$

平均: $\mu$

標準偏差: $\sigma$

$$Z = \frac{X - \mu}{\sigma}$$

**確率変数の標準化**

$$X = \mu + \sigma Z$$

調べやすい

$Z \sim \mathrm{N}(0, 1^2)$

平均: 0

標準偏差: 1

標準正規分布の様子は，教科書巻末の正規分布表を利用して調べる．

## 例題 3: 変数の標準化

ある高校における男子の身長の分布は，平均 170.2cm，標準偏差 5.6cm の正規分布に近いとする．180cm の身長は，この高校で高い方から何 % ぐらいの位置にあるか．

男子の身長を $X$cm とすると，$X \sim \mathrm{N}(170.2, 5.6^2)$.

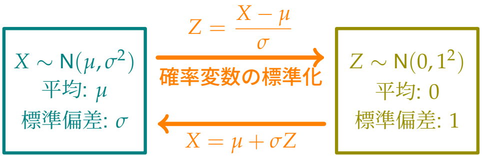
$Z = \frac{X-170.2}{5.6}$ とすると，$Z \sim \mathrm{N}(0, 1)$.

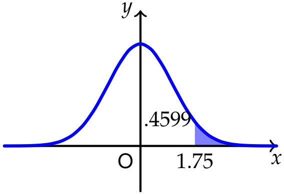
$X = 180$ のとき，$Z = \frac{180-170.2}{5.6} = 1.75$.

$$P(X \geqq 180) = P(Z \geqq 1.75)$$
$$= 0.5 - 0.4599 = 0.0401$$

より，高い方から **4%** ぐらいの位置．

※ 標準正規分布表では，$P(Z \leqq u)$（$u \geqq 0$）の形の確率が載っているだけ，それ以外の区間の確率は，グラフの対称性や $P(Z \geqq 0) = 0.5$ であることを利用する．



# 二項分布と正規分布

正規分布の偉いところは，十分な大きさの標本の標本平均の分布の様子が正規分布に近くなるところにある（このことの詳細な文意は後述）．

例として，二項分布 $B(n,p)$ を取り上げる．二項分布と正規分布の比較のノートブック（`112_BinVsNor.ipynb`）から，標本数 $n$ が大きければ，平均と標準偏差を同じくする正規分布に近づく．

十分大きい $n$ における $B(n,p)$ を，正規分布 $N(np, np(1-p))$ に近似して考え，さらに変数を標準化することで，標準正規分布の確率計算に落とし込む．

**二項分布 $B(n,p)$**

- 平均: $\mu = np$
- 標準偏差: $\sigma = \sqrt{np(1-p)}$

**正規分布の標準化**

$$Z = \frac{X-\mu}{\sigma}$$

**正規分布による二項分布の近似**

$n$ が十分大きいとき $X \sim B(n,p)$ に対して $Z = \dfrac{X-np}{\sqrt{np(1-p)}}$ はほぼ正規分布に従う．

## 例題 4: 正規分布による二項分布の近似

さいころを 50 回投げるとき，1 の目の出る回数 $X$ が次の範囲にある確率を求めよ:

$$\left| \frac{X}{50} - \frac{1}{6} \right| \leqq 0.1$$

## 例題 4: 正規分布による二項分布の近似

さいころを 50 回投げるとき，1 の目の出る回数 $X$ が次の範囲にある確率を求めよ:

$$\left| \frac{X}{50} - \frac{1}{6} \right| \leqq 0.1$$

$X \sim B\left(50, \frac{1}{6}\right)$ より，$Z = \dfrac{X - 50 \cdot \frac{1}{6}}{\sqrt{50 \cdot \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right)}} = \dfrac{6X - 50}{5\sqrt{10}}$ とおくと，$Z \sim N(0, 1^2)$ とみなせる．

$$\left| \frac{X}{50} - \frac{1}{6} \right| \leqq 0.1 \quad \Leftrightarrow \quad |6X - 50| \leqq 30$$

より，与えられた条件は $Z$ の範囲として $|Z| \leqq \frac{30}{5\sqrt{10}} \fallingdotseq 1.90$．よって，求める確率は

$$P(|Z| \leqq 1.90) = 0.4713 \times 2 = \mathbf{0.9426}$$

※ 正規分布による近似を利用せずに計算してみると，$\sum_{|6k-50| \leqq 30} {}_{50}\mathrm{C}_k \left(\frac{1}{6}\right)^k \left(1 - \frac{1}{6}\right)^{50-k} \fallingdotseq 0.9455$ 程度となった．

# 大数の法則と中心極限定理

ノートブック（`112_BinVsNor.ipynb`）において $n$ を大きい数に設定すると，正規分布の山が中央に集まってくるように見える．これは，標本数を大きくすると，標本平均が母平均から大きく外れることは，ほぼなくなることを意味する．この傾向は，正規分布に限らず様々な分布について言えることが古くから知られている．

## 大数の法則（大数の弱法則）

$X_k$（$k = 1, \ldots, n$）は独立で同一分布に従い，その平均 $\mu$ と分散 $\sigma^2$ が存在するとき，任意の整数 $\epsilon > 0$ に対して

$$\lim_{n \to \infty} P\left(\left|\overline{X}_n - \mu\right| \geqq \epsilon\right) = 0$$

※ 証明には，Chebishev の不等式 $P\left(|X - \mu| \geqq \epsilon\right) \leqq \frac{\sigma^2}{\epsilon^2}$ を利用する．分布に関する仮定を様々に敷けば，Chebishev の不等式は自力で証明できるかもしれない．余力があれば，試みられたい．

「二項分布と正規分布」のスライドで，二項分布 $B(n, p)$ は

$n$ が大きければ，平均と標準偏差を同じくする正規分布に近づく

と書いた．一般的に，次の定理が知られている:

## 中心極限定理

平均 $\mu$，分散 $\sigma^2$ をもつ分布からの無作為標本の標本平均 $\overline{X}_n$ の標準化 $Z_n$ の分布は標準正規分布に分布収束する:

$$\lim_{n\to\infty} P(Z_n \leqq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

※ この定理は現段階で扱える数学の程度を大幅に超える．式の意味を触れるに留める．
左辺の $P(Z_n \leqq u)$ は $Z_n \leqq u$ となる確率であるから，標準化された確率変数 $Z_n$ の分布関数（変数は $u$）を表している．右辺の定積分の被積分関数は，標準正規分布の確率密度関数である．その区間 $(-\infty, u)$ における定積分であるから，右辺は標準正規分布の分布関数を表している．

中心極限定理は，左辺の分布関数において，標本数 $n$ を限りなく大きくしたとき，標準正規分布の分布関数に近づくことを主張している．