高校数学における平均と中心極限定理について

中央大学　名誉教授、統計科学研究所
杉山　高一
大阪府立大学　生命環境科学域　数理科学課程
兵頭　昌

1．平均の説明と平均は何故重要か
2．正規分布と標本平均の分布
3．正規分布の易しい導入
4．平均の信頼幅
5．母集団分布が二項分布の場合

要約：母比率 pの近似信頼区間の導出を概説することに重点を置いている. 高校教育における区間推定において、中心極限定理がその導出法にどのように応用されているかを具体的に示す.

1.平均の説明と平均はなぜ重要か

推測統計の考え

統計学では母集団に正規分布や2項分布などの確率分布を想定する．これを**母集団分布**と呼ぶ．一般に，確率分布は未知の母数(パラメータ)を含むから，母集団からの無作為標本に基づいて，未知の母数を推測することになる．これを**統計的推測**と呼ぶ．例えば，母集団に正規分布 $\mathcal{N}(\mu, \sigma^2)$ を仮定すると，正規分布は2つのパラメータ μ, σ^2 を含むから，これらのパラメータを標本に基づいて推測することにより母集団分布が規定され，母集団の性質・特徴を把握することが可能となる．

統計量と標本分布について

パラメータに関する推測は，無作為標本 $X_1, X_2, \cdots, X_n$ の関数を用いて行われる．これを**統計量**という．例えば，

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

$$S^2 = \frac{1}{n-1}\{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\}$$

は統計量であり，それぞれ**標本平均**，**標本分散**と呼ばれる．また，S を**標本標準偏差**という．標本平均と標本分散は，母集団分布の平均 (母平均) μ と分散 (母分散) σ^2 を推測するのに用いられる．母集団分布のパラメータを推測する際に，そのパラメータに対応する統計量の分布を知ることが必要になってくる．統計量の変動の様子を表す分布を**標本分布**と呼ぶ．

標本平均について

$\mathrm{E}[X]=\mu, \mathrm{Var}[X]=\sigma^2$ である母集団分布より，無作為標本 $X_1, X_2, \cdots, X_n$ が得られたとする. このとき,

$$\mathrm{E}[\bar{X}]=\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[X_i\right]=\mu$$

となる. これを, **不偏性**という.

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

大標本における良い性質 1

標本の大きさ n が大きくなると $\bar{X}$ の分散は小さくなり，分布は μ のまわりに集中してくることがわかる．このことは，n が大きいほど μ のより精確な推測が可能なことを意味している.

標本平均について

大標本における良い性質 2

標本の大きさ n が大きくなると $\bar{X}$ の μ 周りの分布は, 適当な条件の下で正規分布に近づく.

⇓

μ の区間推定へ利用する.

この事実は, 母集団分布へ母比率 p のベルヌーイ分布を想定しても成立する.

⇓

比率 p の区間推定へ利用する.

2.正規分布と標本平均の分布

標本平均の分布

目的：今年の18歳の女性の身長の平均μを知る

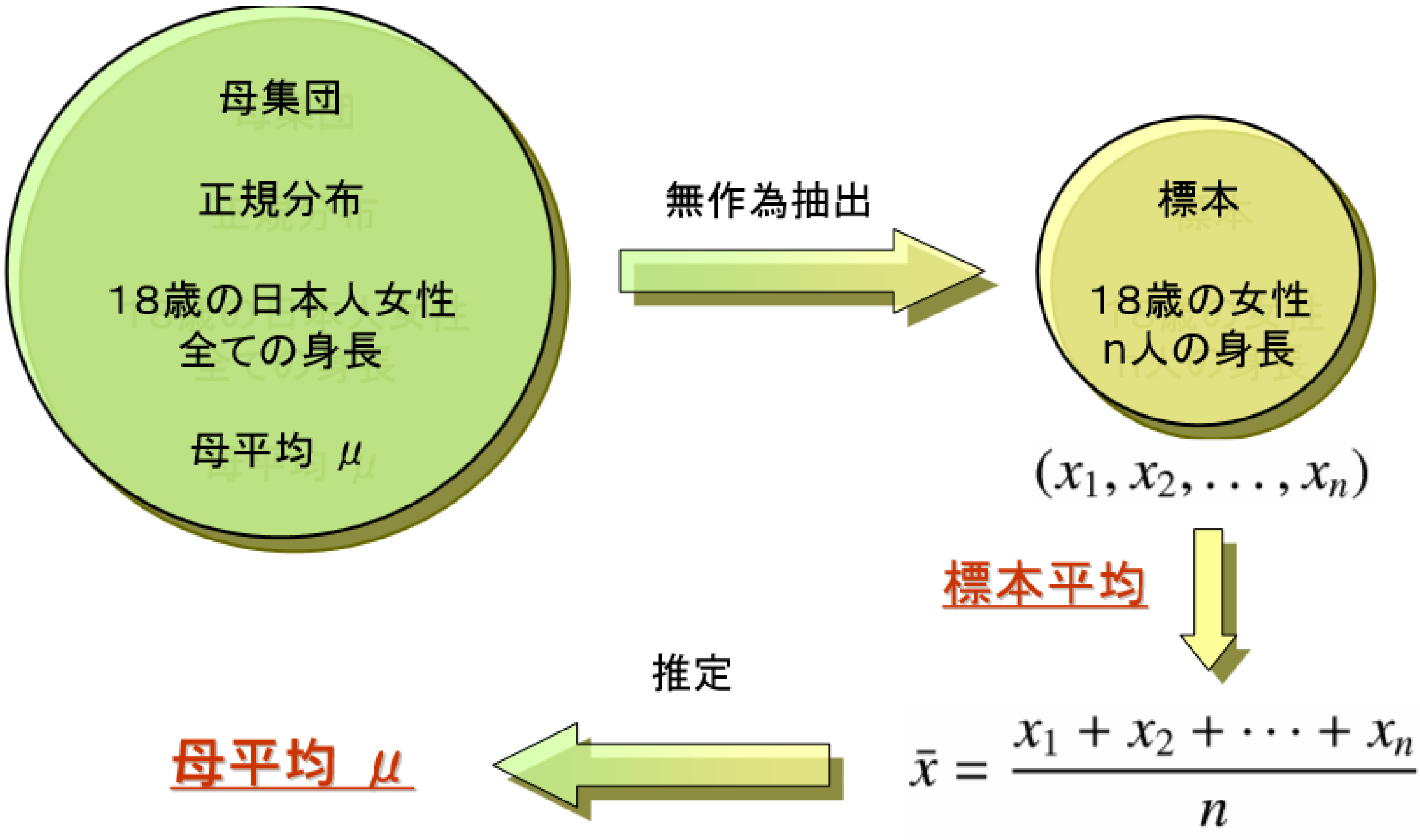
- 方法1：全ての18歳の女性の身長を調査
  ⇒ 不可能に近い
- 方法2：母集団から抽出した標本に基づいてμを推定
  ⇒ 現実的な方法（標本平均による推定）

例：抽出する標本の大きさを5とする

- 標本：159，149，163，154，160
- 標本平均：

$$\bar{x} = \frac{159 + 149 + 163 + 154 + 160}{5} = 157$$

標本平均による母平均の推定



抽出する標本によって，標本平均の値がばらつく

標本平均の分布

実験：5人を無作為抽出し、身長の平均を求める

- 1回目の実験の5人から平均を計算

  157.1, 157.3, 158.7, 161.3, 155.3

  標本平均 = 157.9

- 2回目の実験の5人から平均を計算

  160.1, 157.6, 156.1, 153.4, 157.0

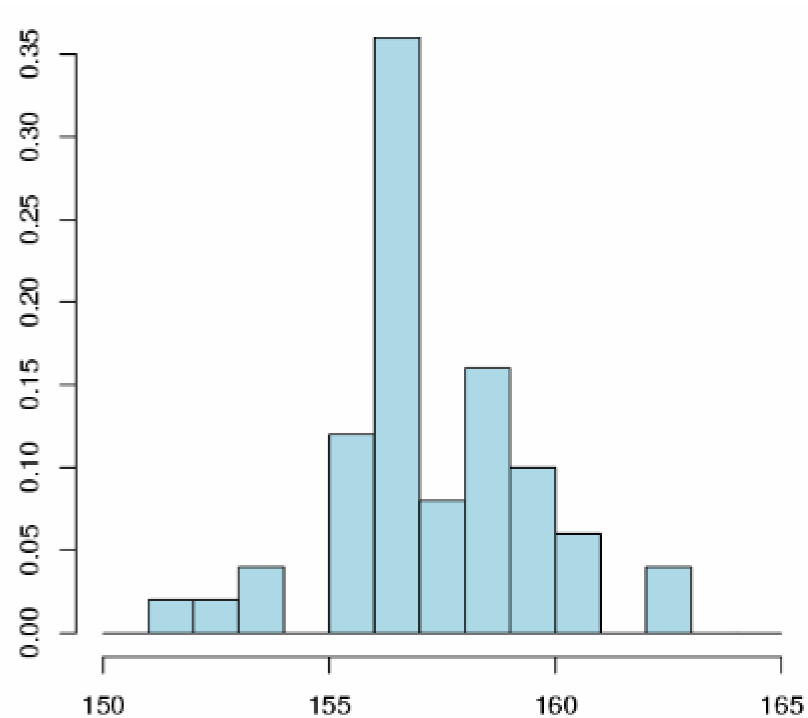
  標本平均 = 156.8

⋮

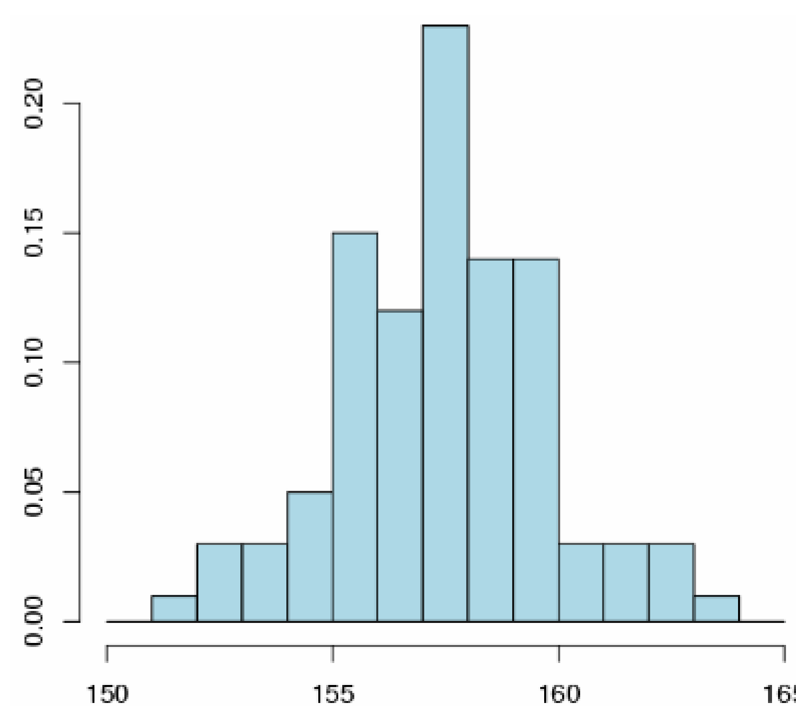
実験を繰り返し行い，ヒストグラムを描く
（標本平均の分布についての考察）

標本平均のヒストグラム

50回の抽出実験によるヒストグラム



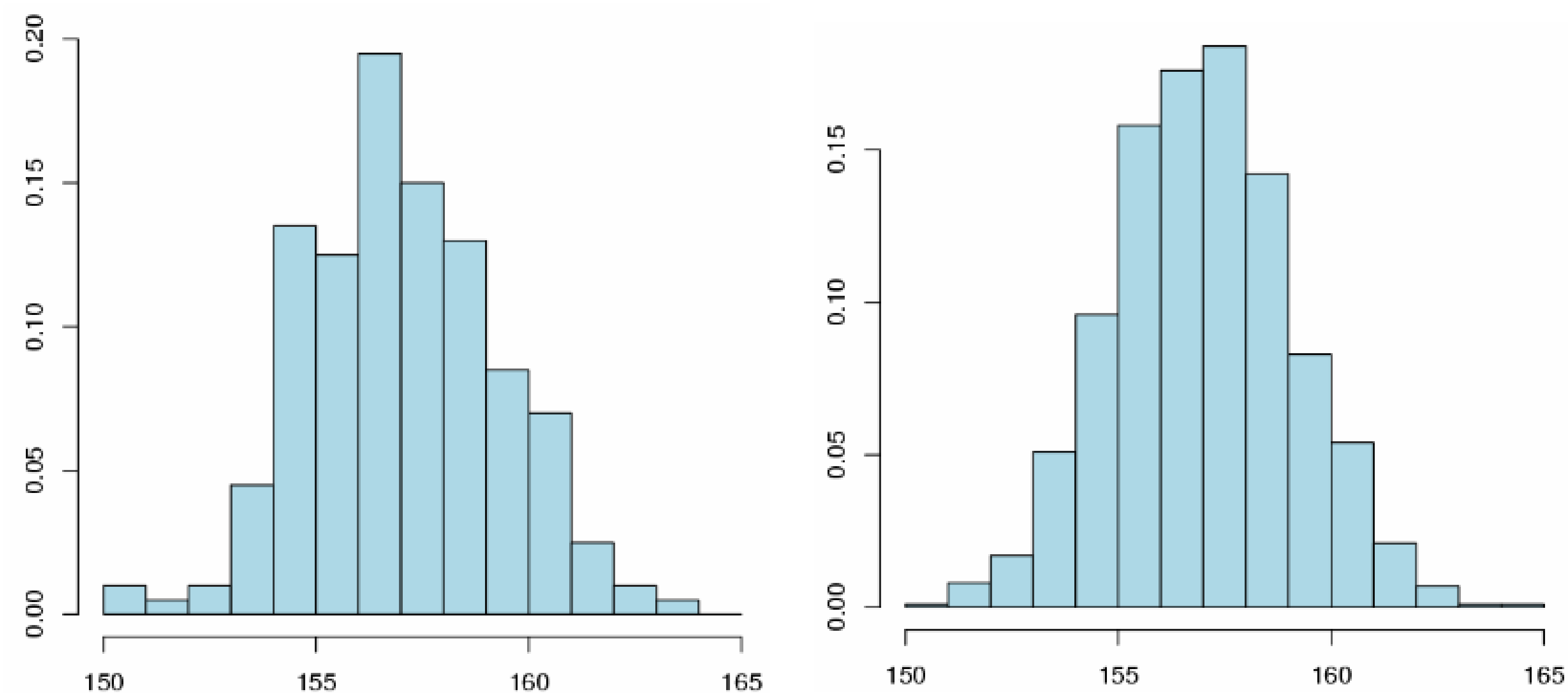
100回の抽出実験によるヒストグラム



<u>繰り返し回数が少ないと不安定</u>

標本平均のヒストグラム

200回の抽出実験によるヒストグラム　1000回の抽出実験によるヒストグラム

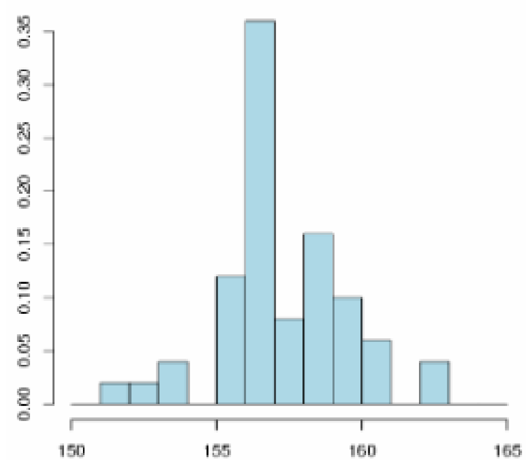


繰り返し回数が多くなるにつれ安定してくる

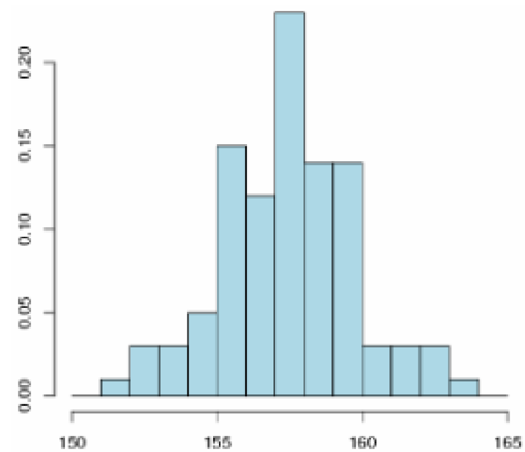
<u>ある正規分布に近づく様子が見られる</u>

標本平均のヒストグラム

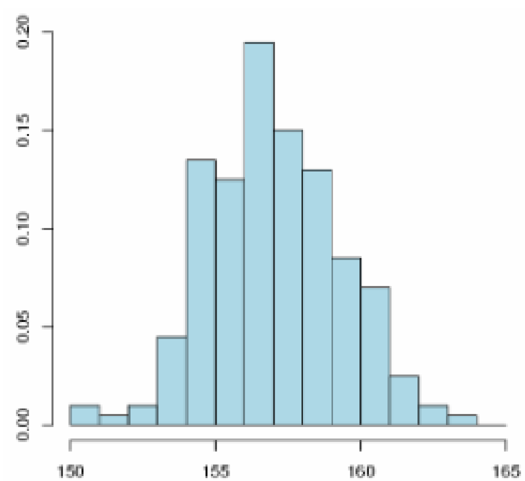
50回



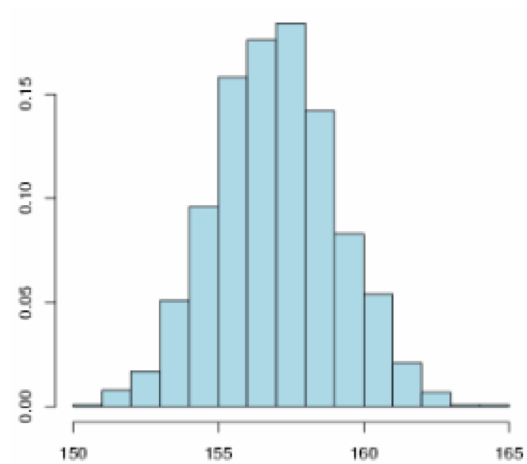
100回



200回



1000回



ある正規分布に近づく

標本平均の分布

平均 μ, 分散 σ^2 の正規分布にしたがう母集団から抽出された大きさ n の標本にもとづく標本平均

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

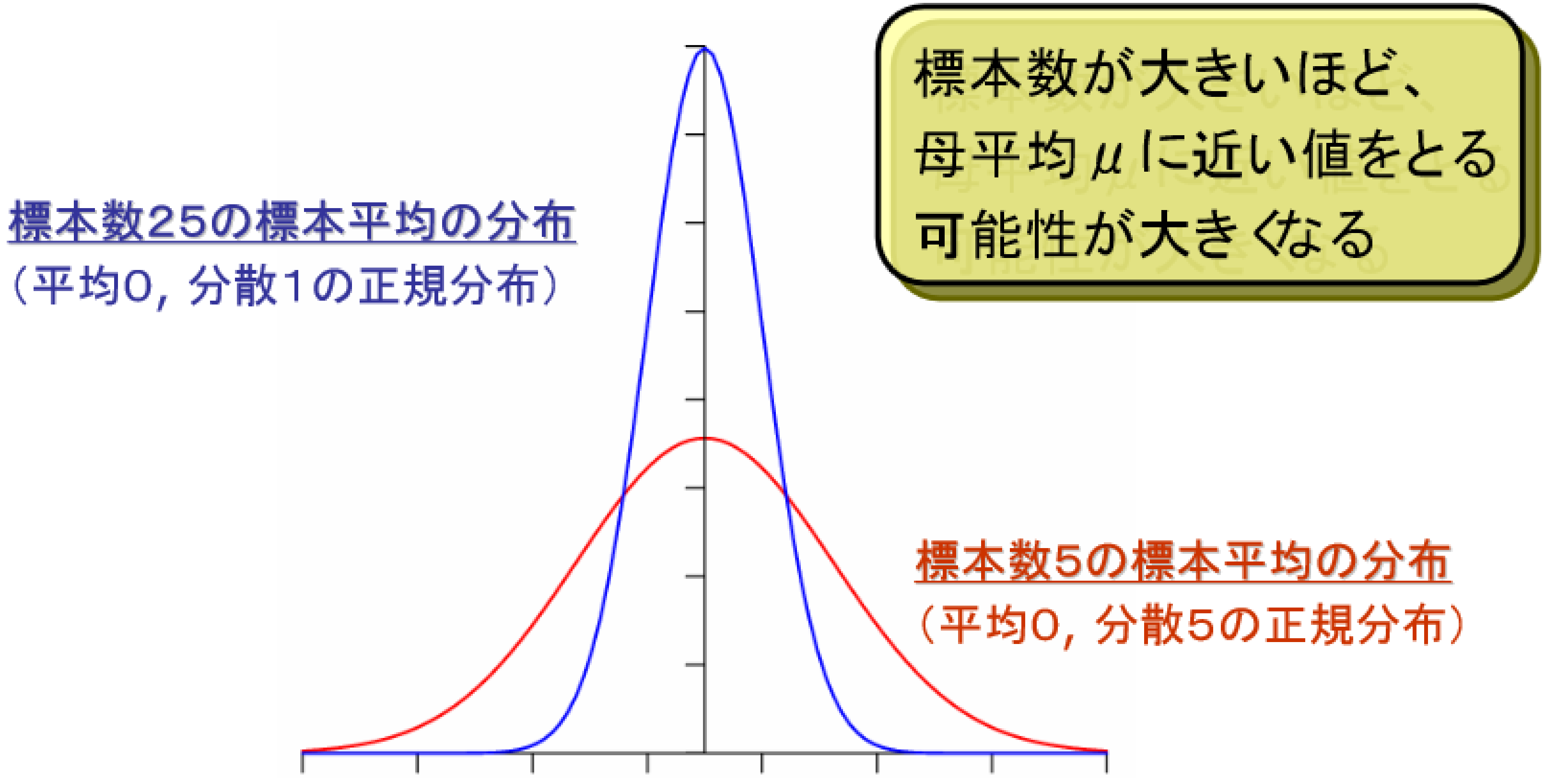
の分布は,

平均 μ, 分散 $\dfrac{\sigma^2}{n}$ の正規分布

にしたがう

標本の大きさと標本平均の分布

- 標本の大きさが5と25の場合の標本平均の分布
- 標本は平均0，分散25の正規分布から抽出



標本平均の分布

変量Xが平均μ，分散σ^2の分布にしたがう
（正規分布とはかぎらない）

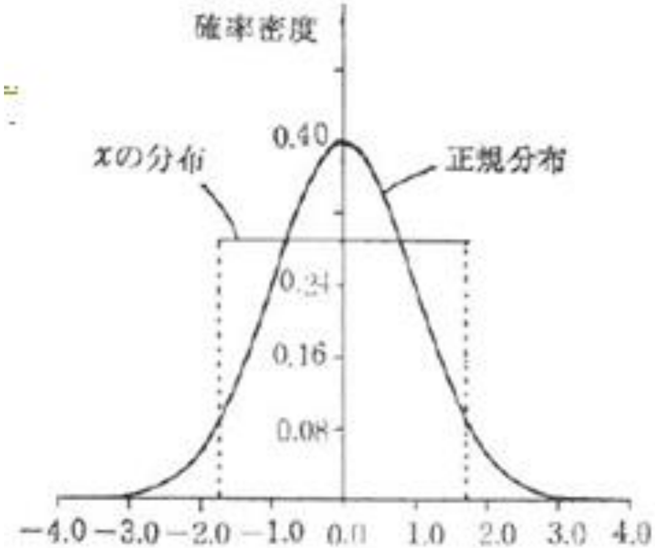
⇩

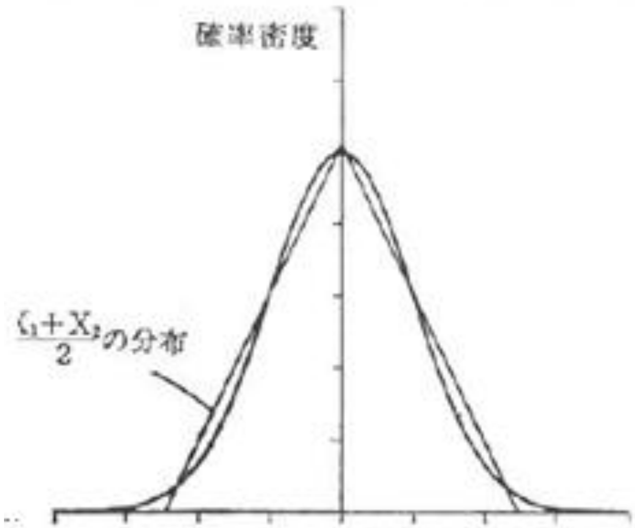
大きさnの標本平均

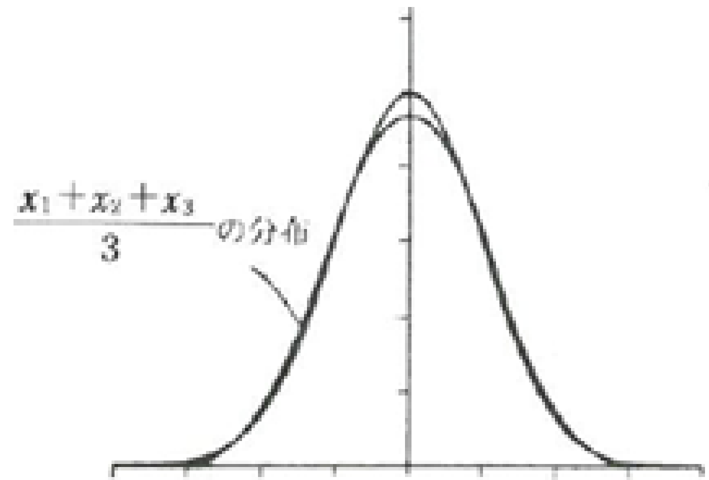
$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

は，近似的に平均μ，分散σ^2/nの正規分布にしたがい，
標本数nが大きくなるにつれてこの近似も良くなる

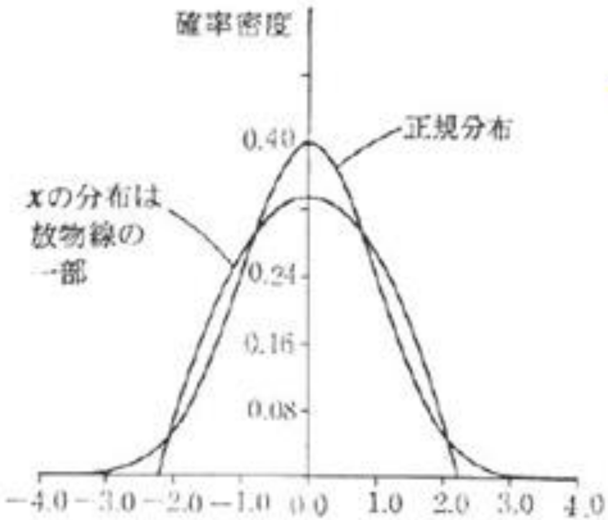
一様分布

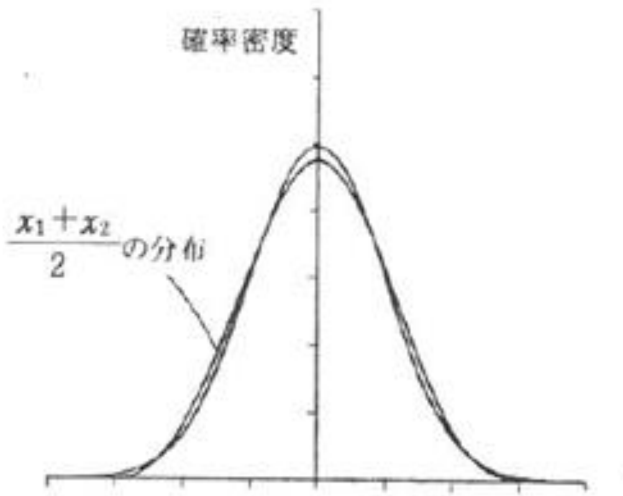

確率密度
0.40
xの分布
正規分布
0.24
0.16
0.08
−4.0 −3.0 −2.0 −1.0 0.0 1.0 2.0 3.0 4.0

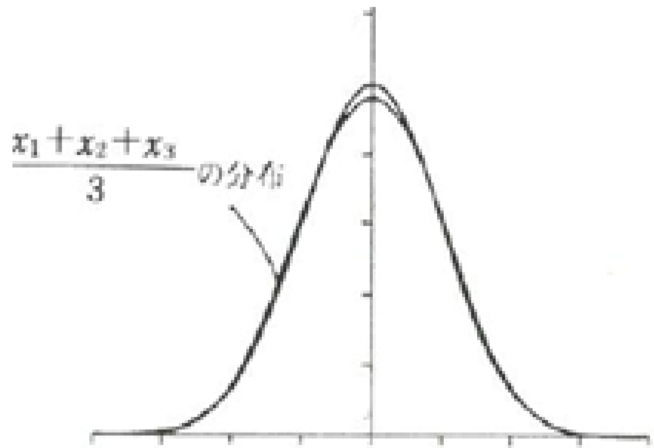

確率密度
$\frac{X_1+X_2}{2}$の分布


$\frac{x_1+x_2+x_3}{3}$の分布

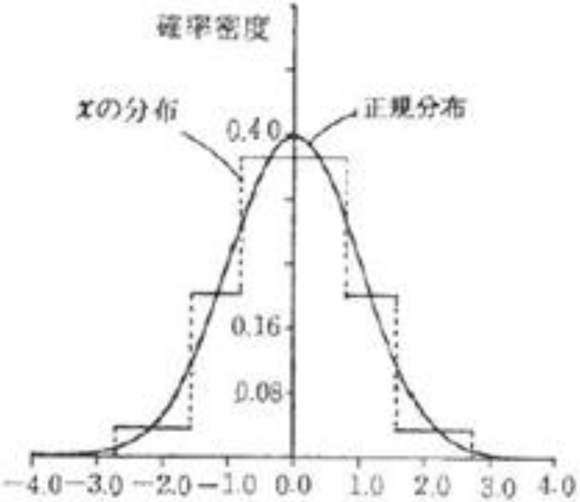
放物線の形状をした分布

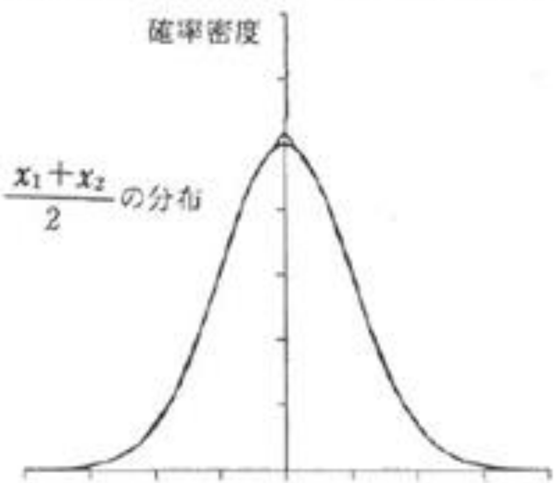


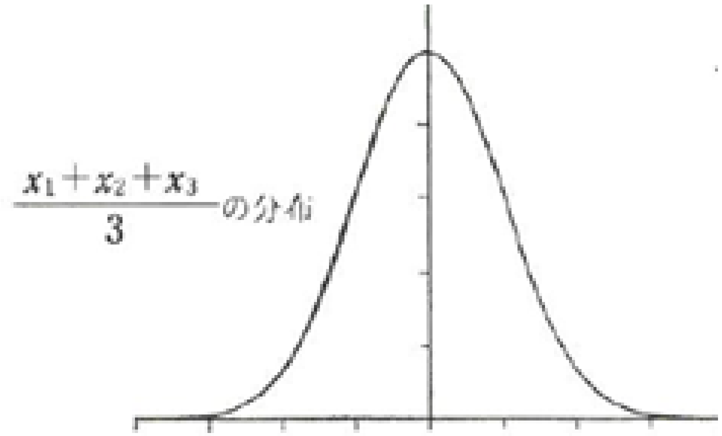




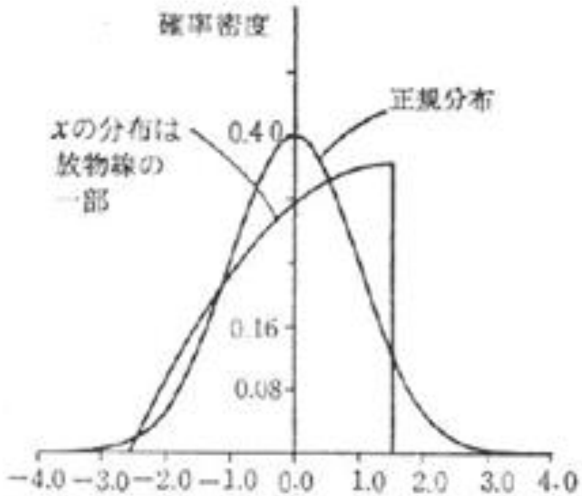
清水先生の導出した分布



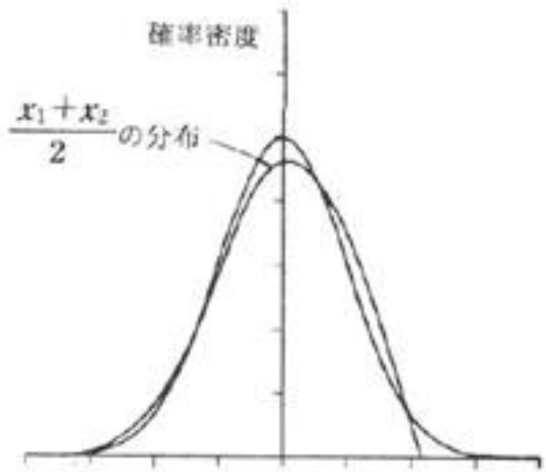




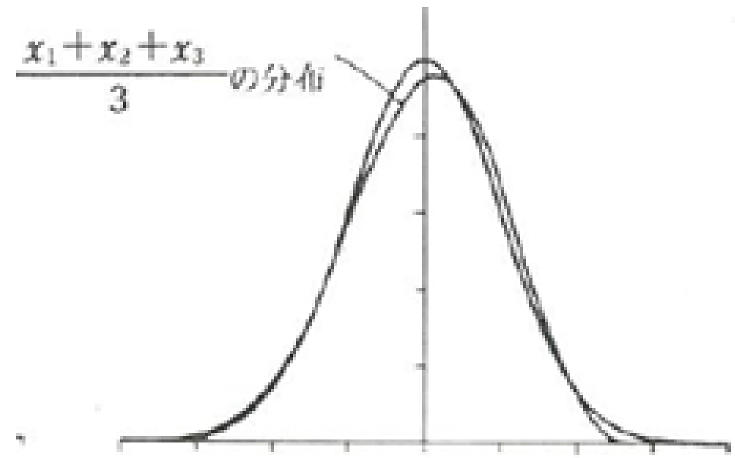
歪んだ分布



確率密度
正規分布
xの分布は放物線の一部
0.40
0.16
0.08
−4.0 −3.0 −2.0 −1.0 0.0 1.0 2.0 3.0 4.0



確率密度
$\frac{x_1+x_2}{2}$の分布



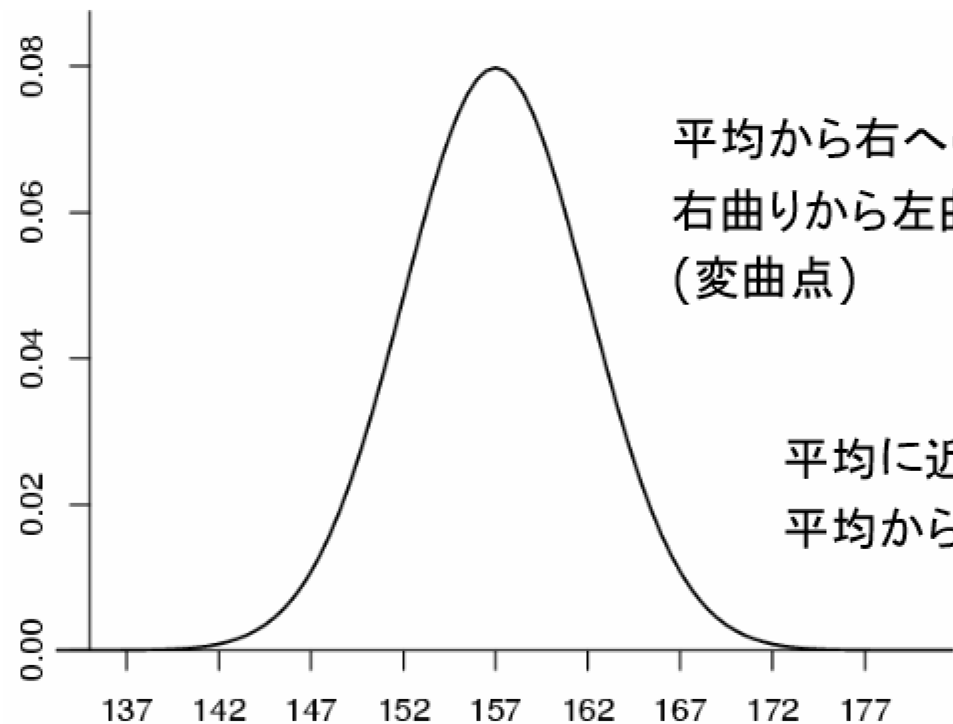
$\frac{x_1+x_2+x_3}{3}$の分布

3.正規分布のやさしい導入

正規分布

平均157, 分散25の正規分布

平均157.0を中心に左右対称なつりがね型



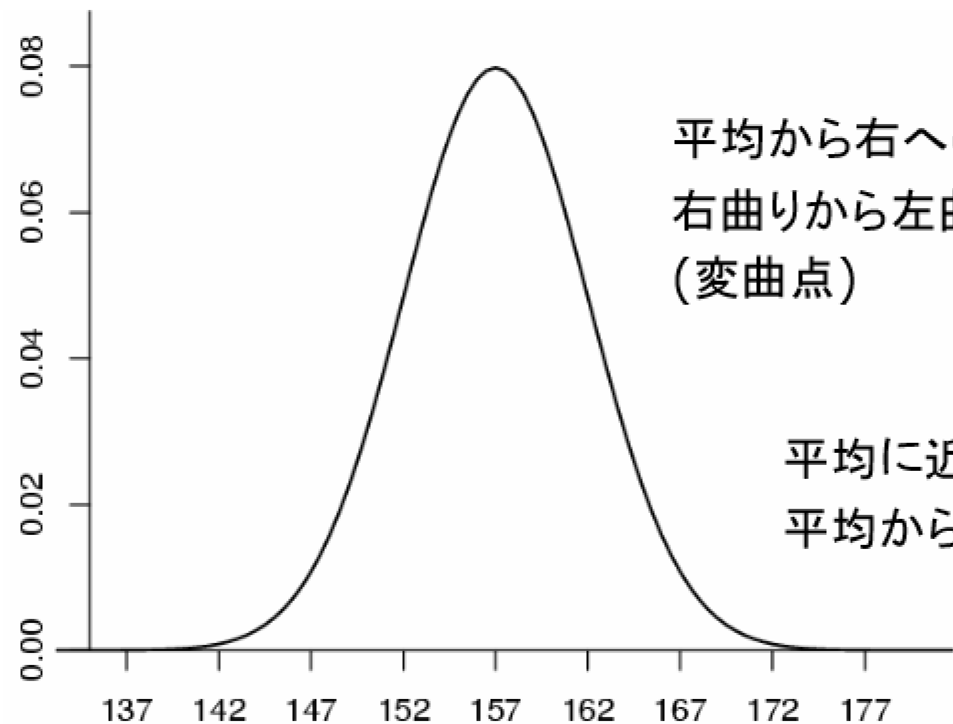
平均から右へ(左へ)5.0のところで
右曲りから左曲りに変わっている
(変曲点)

平均に近いほど高く
平均から遠いほど低い

正規分布

平均157, 分散25の正規分布

平均157.0を中心に左右対称なつりがね型



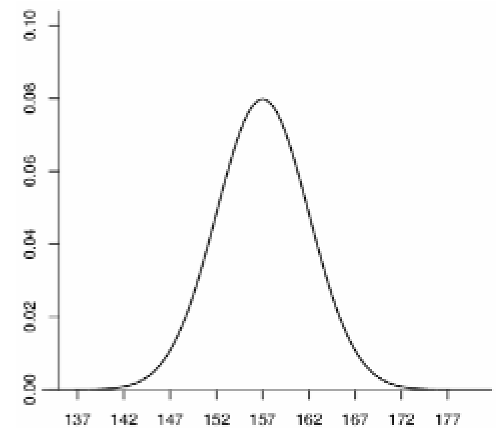
平均から右へ(左へ)5.0のところで
右曲りから左曲りに変わっている
(変曲点)

平均に近いほど高く
平均から遠いほど低い

標準偏差・分散

正規分布

- 生物学的な測定値に見られる
  - 例：身長、座高、腕の長さ
- 測定誤差にもあてはまる



標準偏差と分散

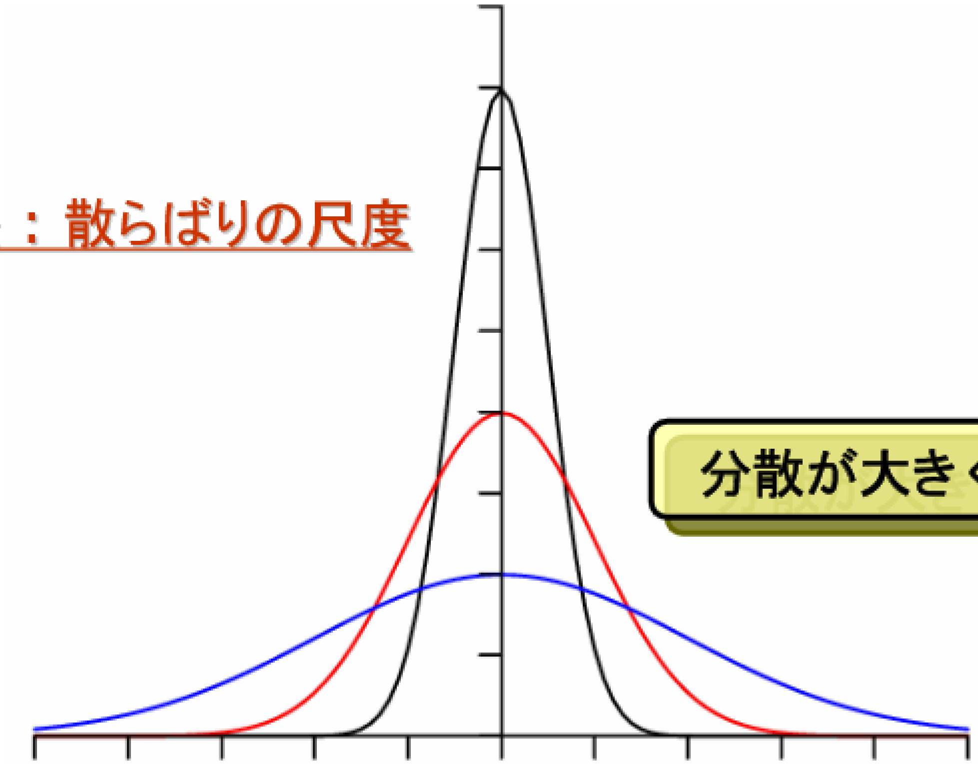
- 標準偏差：平均から変曲点までの距離（σ で表す）
- 分散：標準偏差の平方（σ^2 で表す）
- 標準偏差・分散：散らばりの尺度

分散

- 平均157，σ＝2.5，5，10の場合の正規分布

分散が小さいと平均の周りに集中

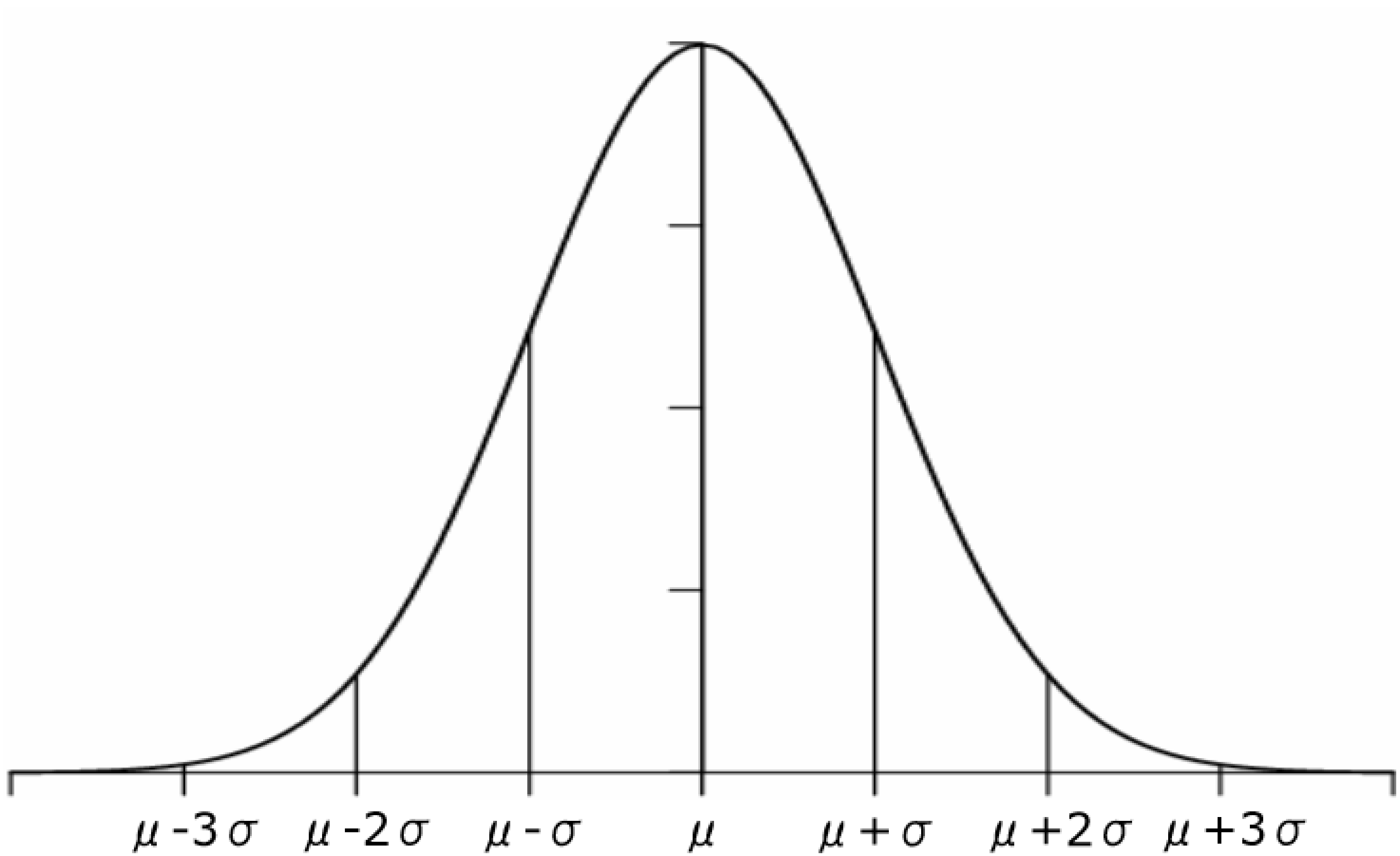
分散、標準偏差：散らばりの尺度



分散が大きくなると裾広がり

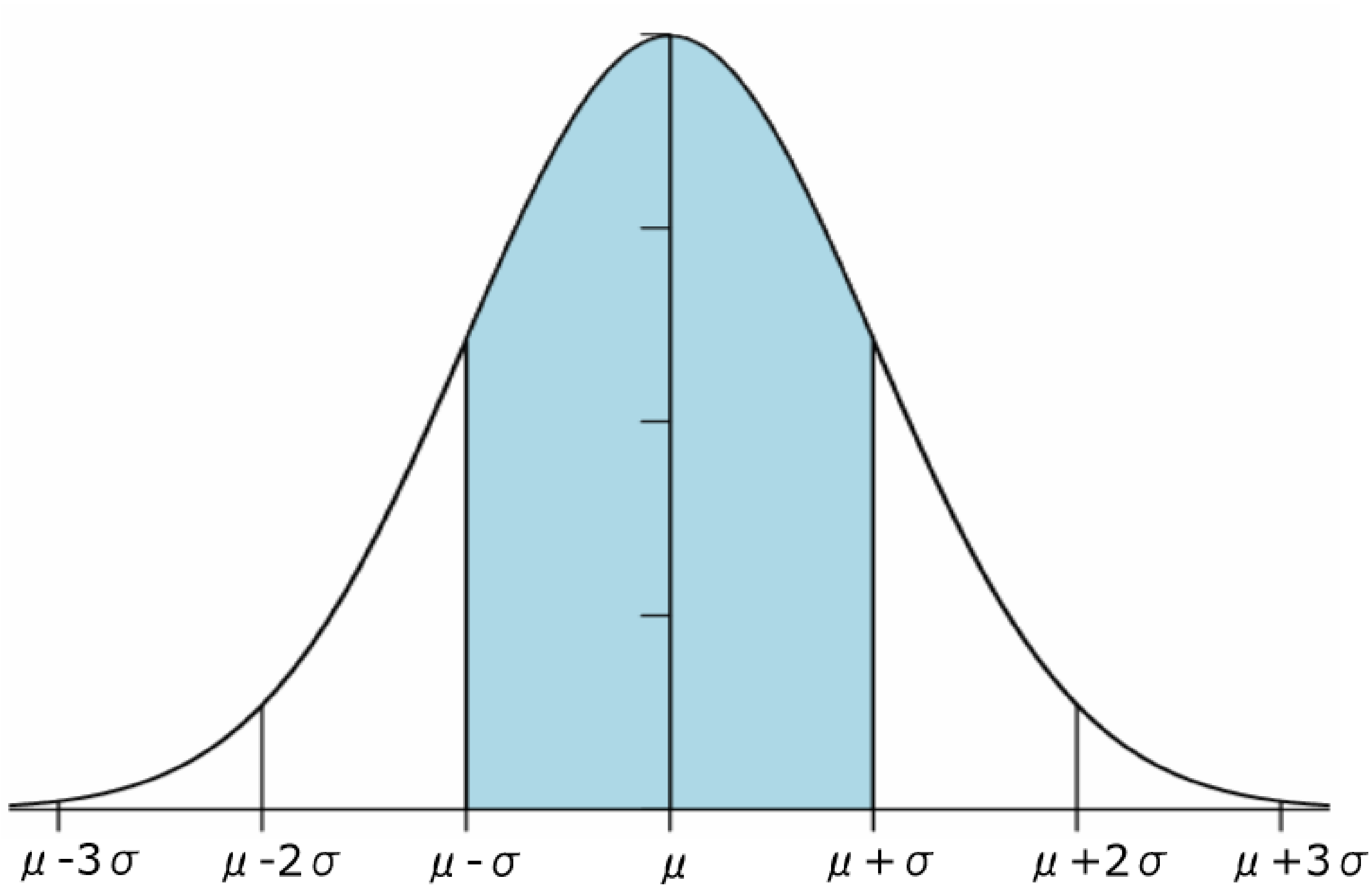
正規分布の性質

正規分布はμとσにより分布が定まる(**母数**)



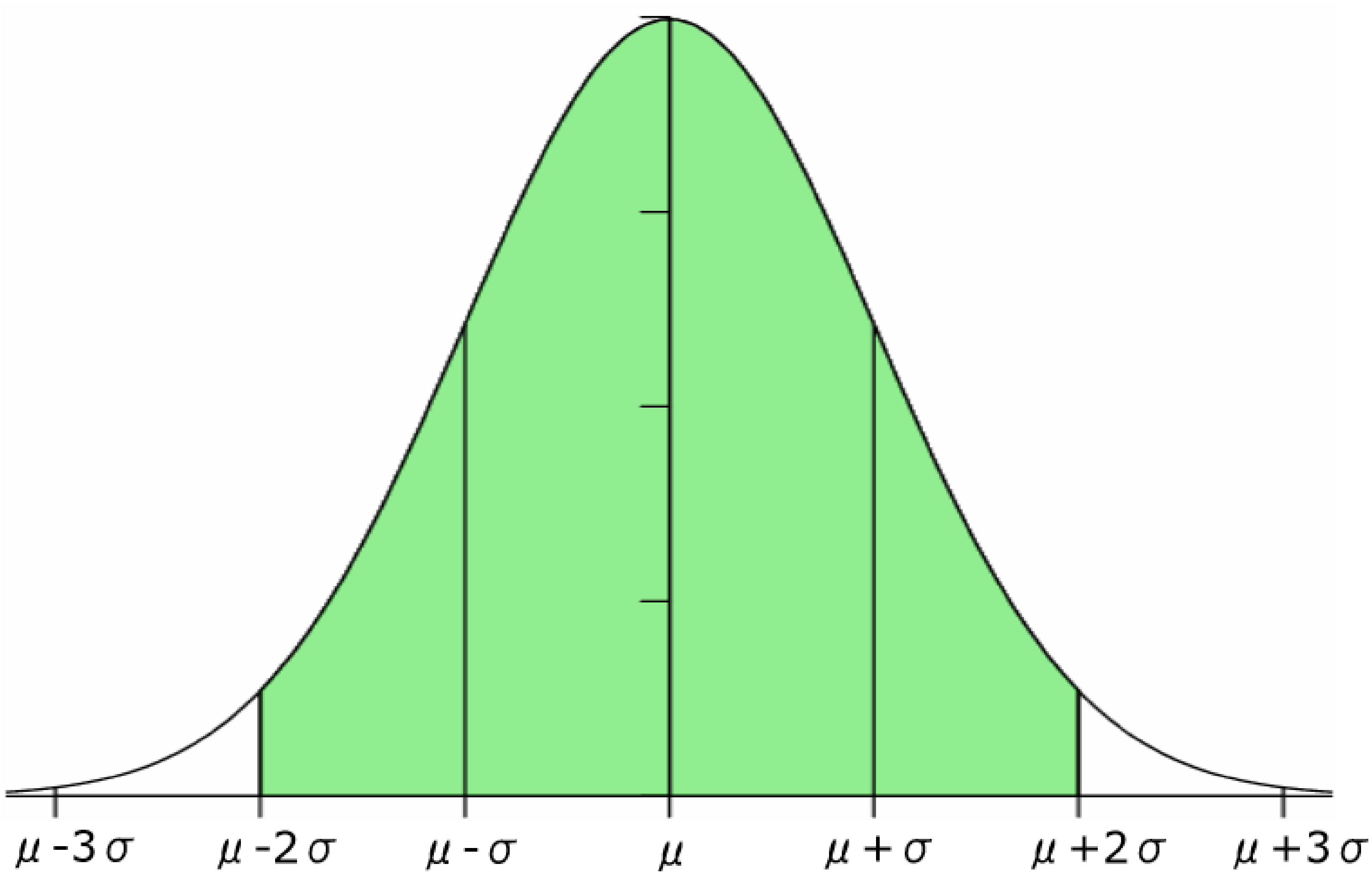
正規分布の曲線下の面積

青で塗りつぶした面積は，全体面積の68.3%



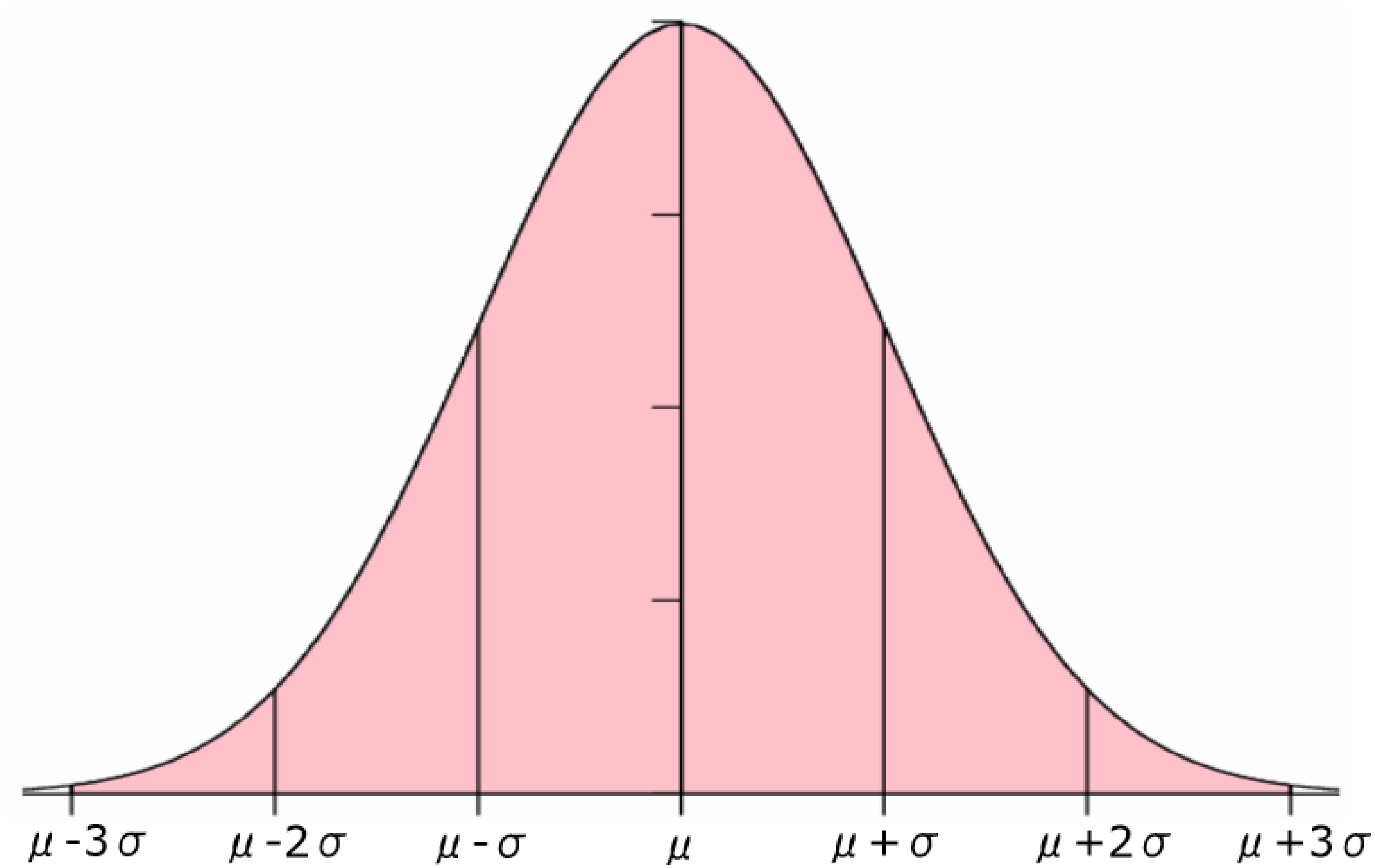
正規分布の曲線下の面積

緑で塗りつぶした面積は，全体面積の95.4%



正規分布曲線下の面積

緑で塗りつぶした面積は, 全体面積の99.7%



離散型･連続型確率変数

離散型確率変数

- 確率変数 x の各値に確率が対応している
  - 例 ： ある夫婦に女児の生れる人数x

連続型確率変数

- 確率変数 x が連続的に変化する（身長 等）
- ある1点の確率は 0

標準正規分布

標準正規分布

- 平均 $\mu = 0$, 分散 $\sigma^2 = 1$ の正規分布

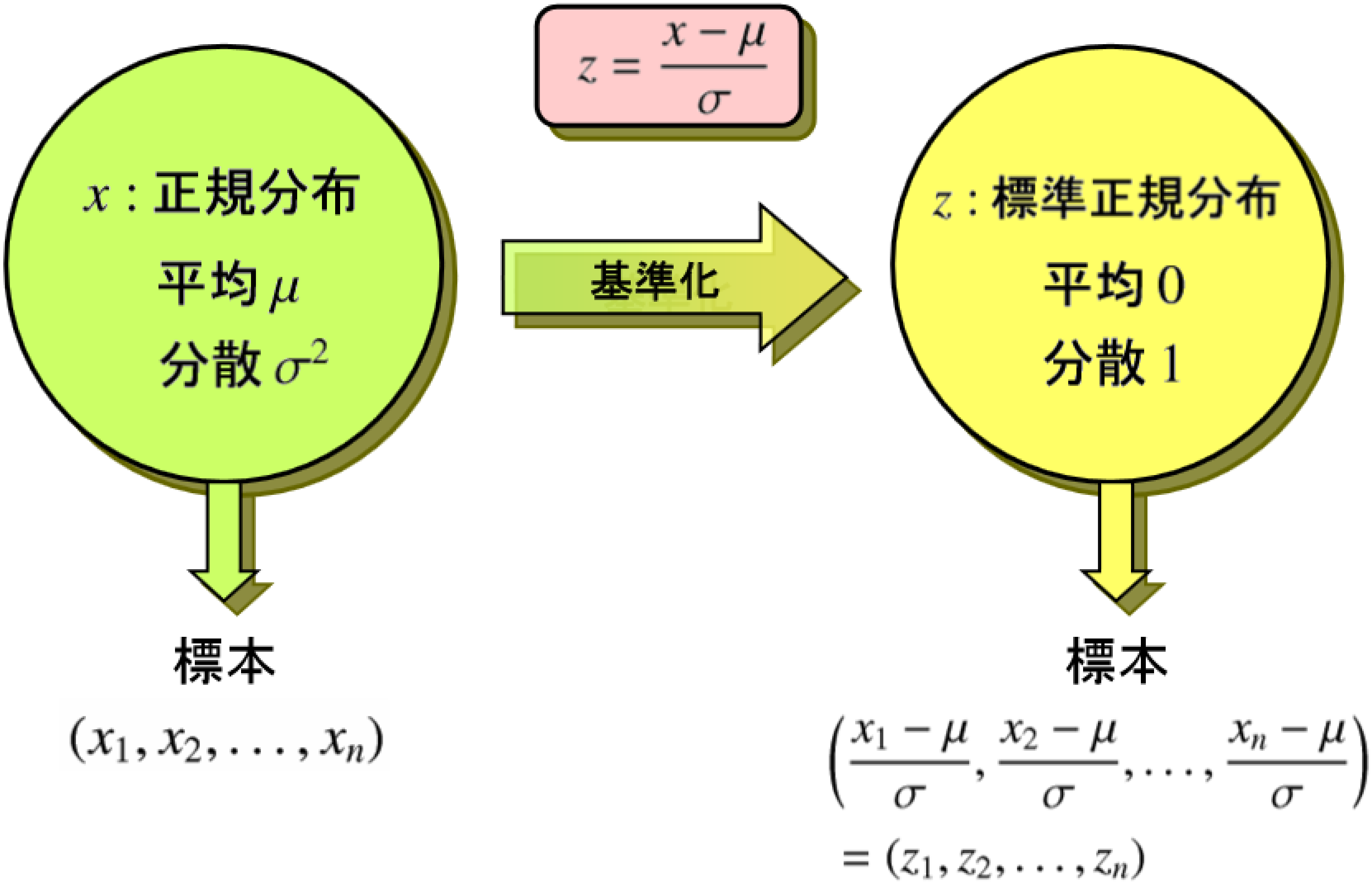
- 基準化
  - 変量 X が平均 μ, 分散 σ^2 の正規分布にしたがうとき,

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

とおくと, z は標準正規分布にしたがう

$$\Pr(X \leq x) = \Pr\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

標準正規分布への変換



$z = \dfrac{x-\mu}{\sigma}$
x : 正規分布
平均 μ
分散 σ^2
基準化
z : 標準正規分布
平均 0
分散 1
標本
$(x_1, x_2, \ldots, x_n)$
標本
$\left(\dfrac{x_1-\mu}{\sigma}, \dfrac{x_2-\mu}{\sigma}, \ldots, \dfrac{x_n-\mu}{\sigma}\right)$
$= (z_1, z_2, \ldots, z_n)$

例

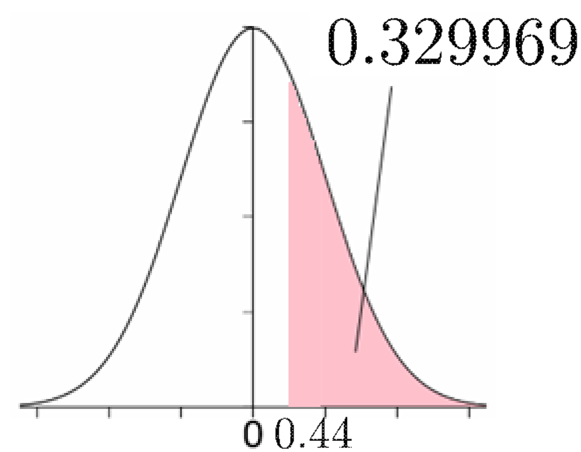
18歳の女性の身長 x

- 平均 $\mu = 157.0$, 分散 $\sigma^2 = 5.0^2$ の正規分布にしたがう

標準正規分布への変換（基準化）

- 変数 x の値が 159.2 であったとする

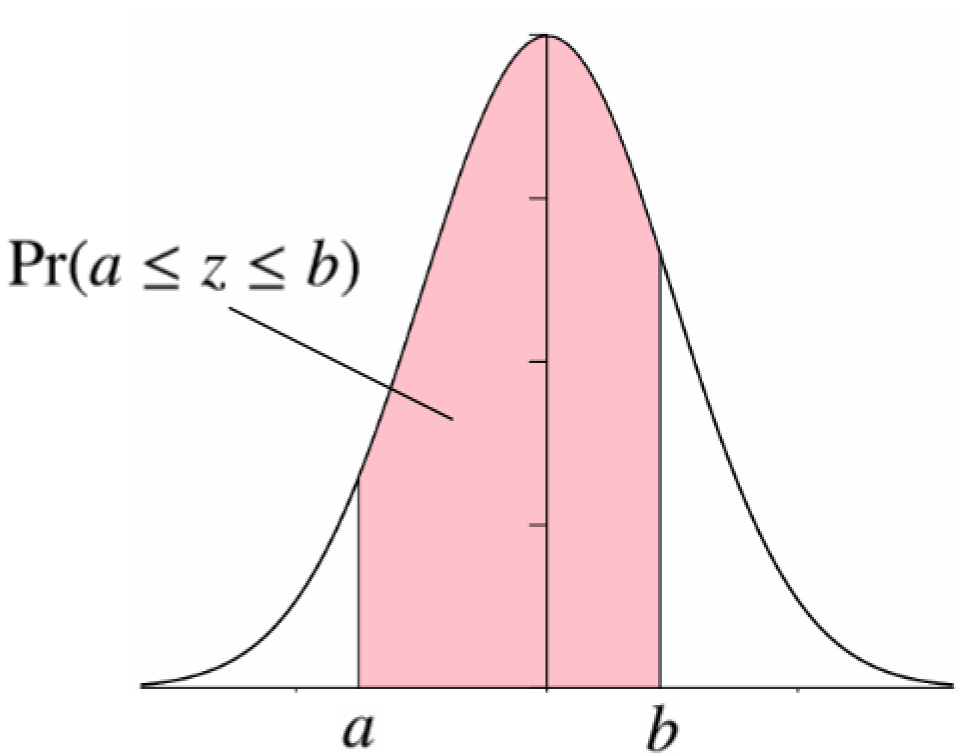
$$z = \frac{x-\mu}{\sigma} = \frac{159.2 - 157.0}{5.0} = 0.44$$



標準正規分布の確率

標準正規分布の確率(ピンク色の部分の面積)

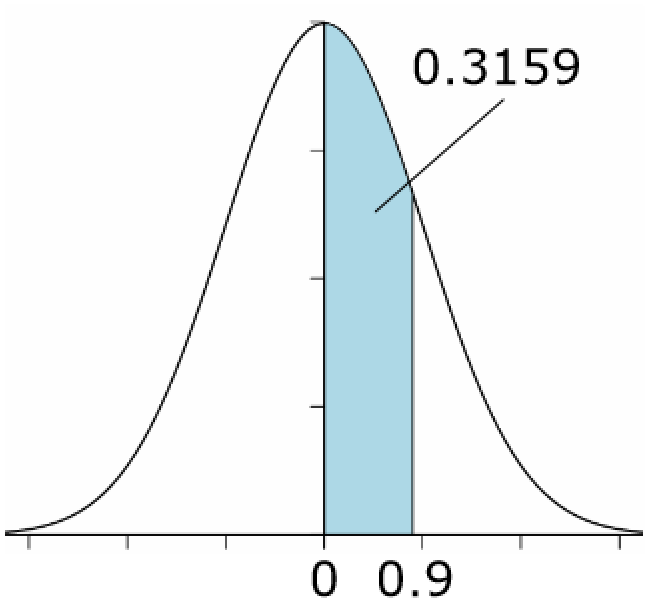
- 参考書等の付表から得られる.
- 統計パッケージの利用により得られる.

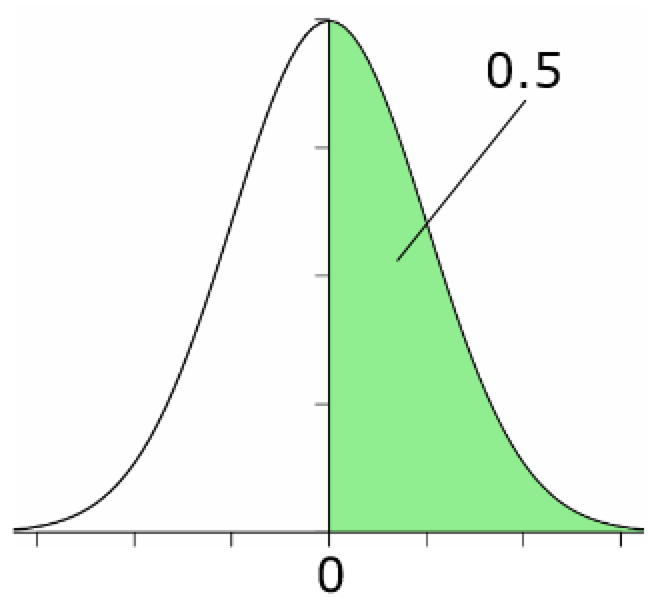


例

標準正規分布の確率

- 下に示した確率は，統計パッケージ及び参考書等の付表より得られる

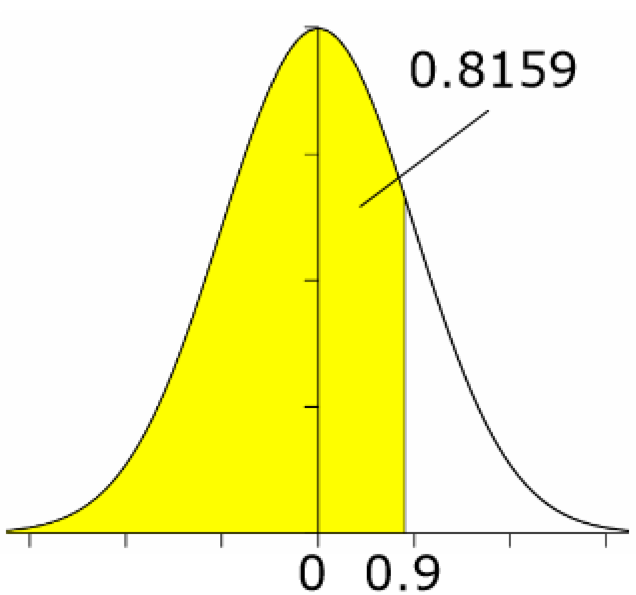


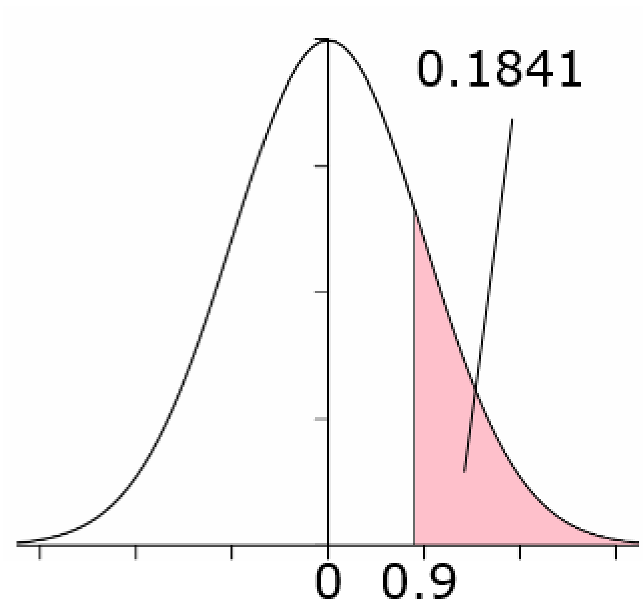


標準正規分布の確率

標準正規分布の確率

- 下に示した確率は，統計パッケージ及び参考書等の付表より得られる

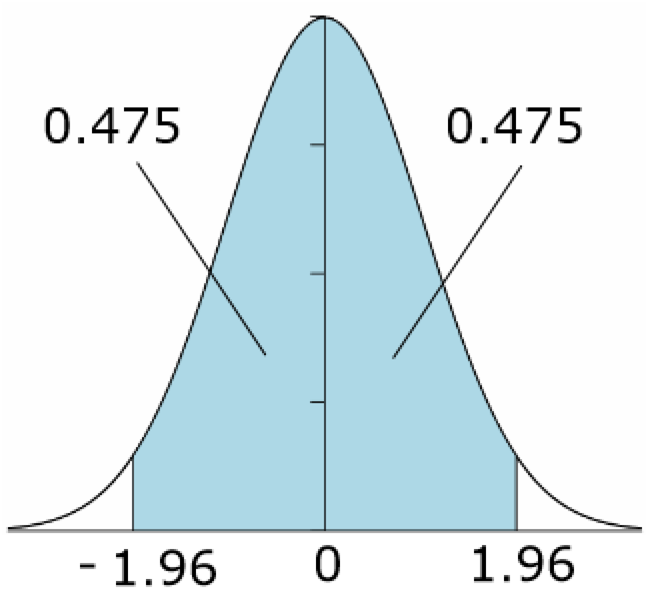


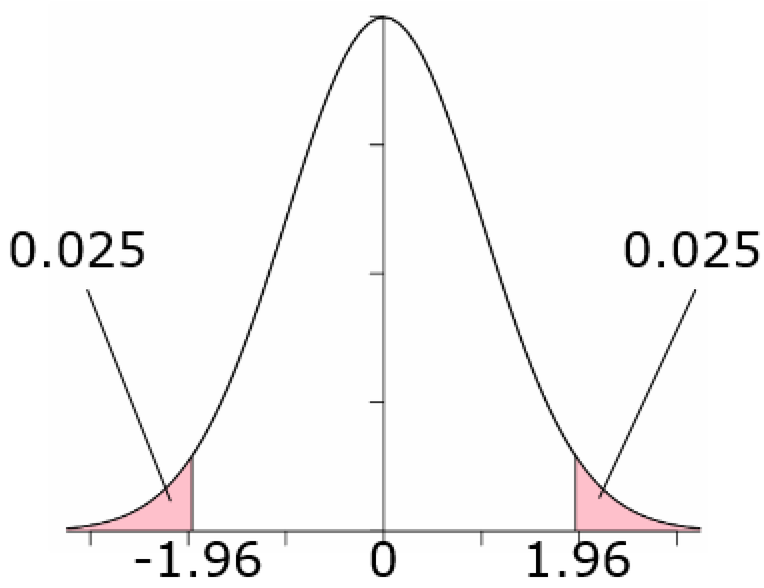


標準正規分布の確率

標準正規分布の確率

- 下に示した確率は，統計パッケージ及び参考書等の付表より得られる

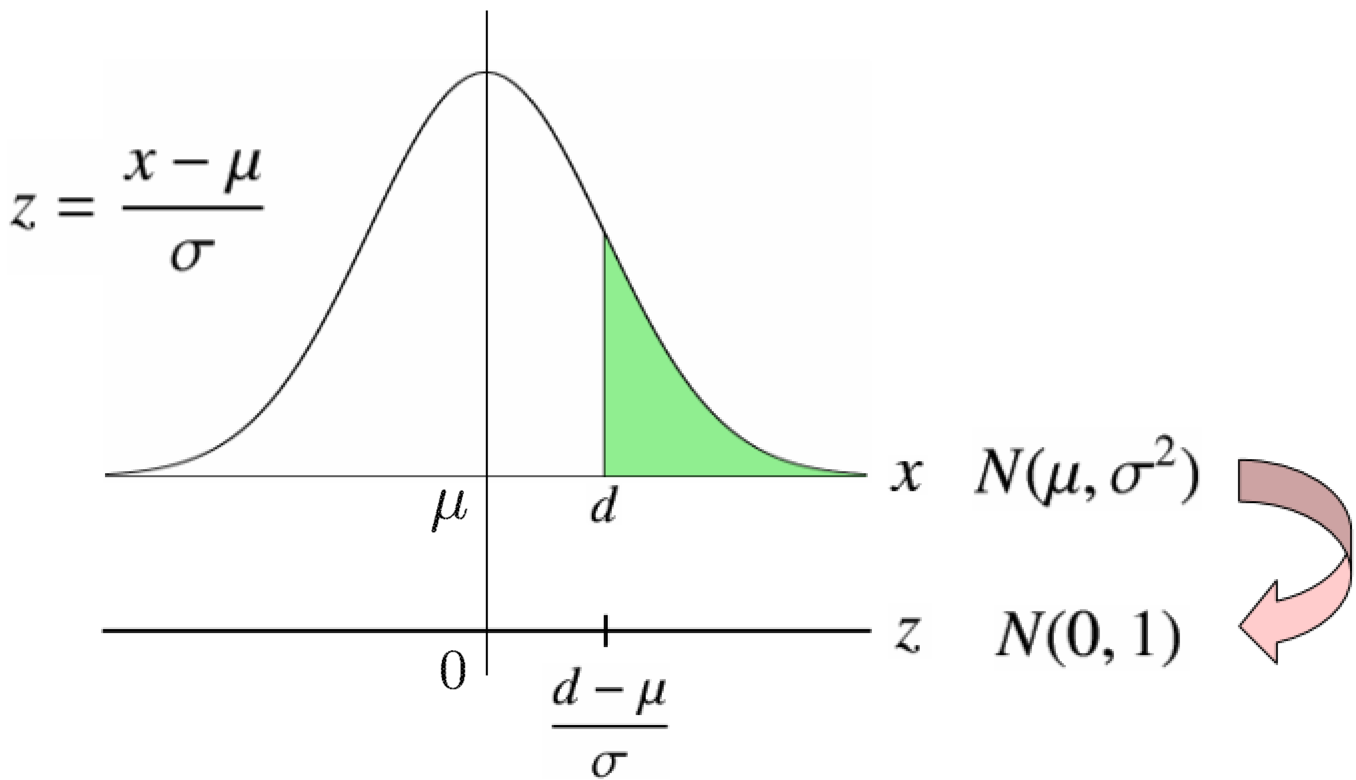




正規分布の確率

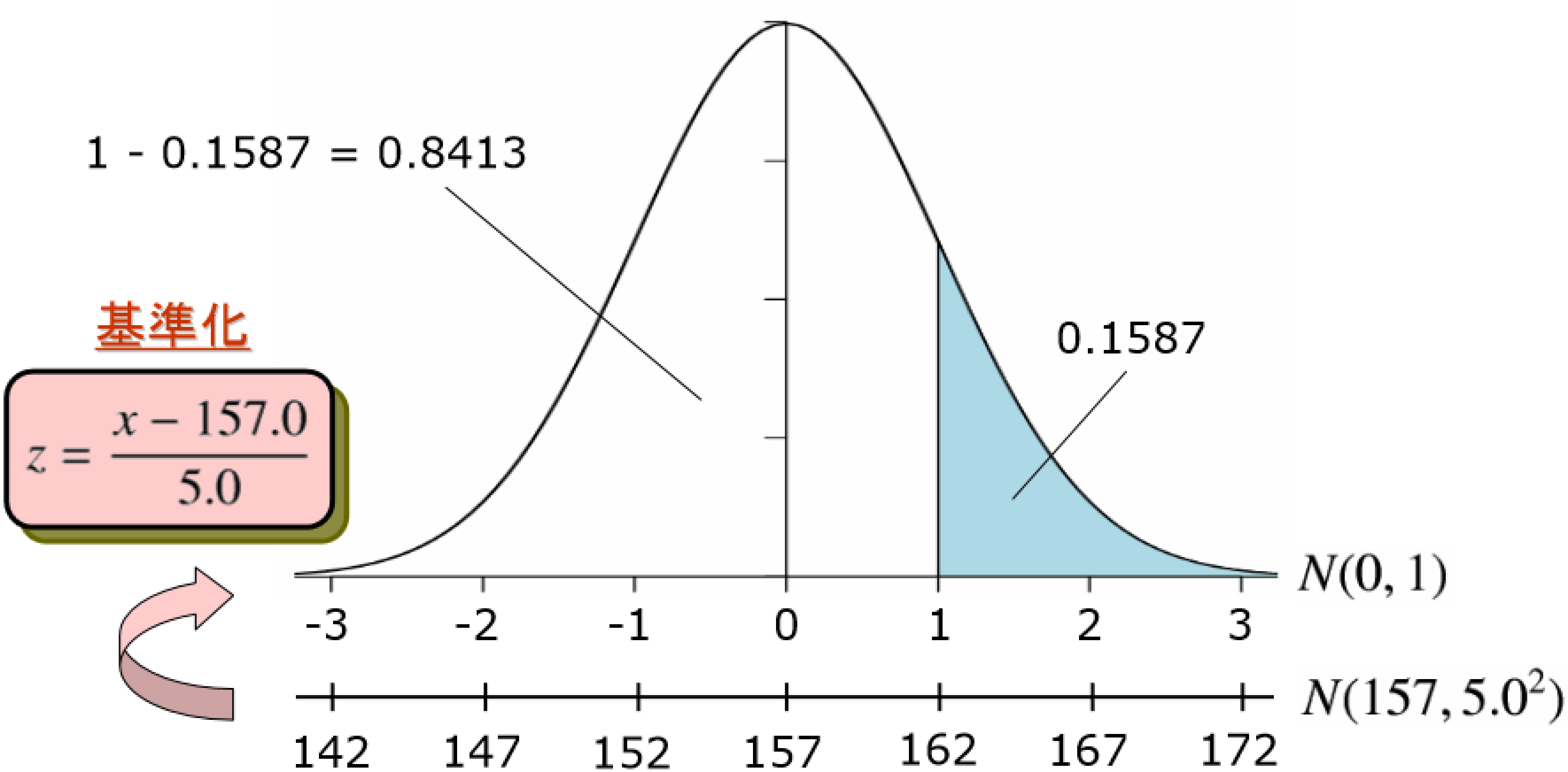
正規分布の確率の求め方

- <u>標準正規分布への変換</u>から求める



例：正規分布の確率

- 平均157.0，分散25の正規分布で，162以上の値をとる確率



４．平均の信頼幅

推定

- **点推定**
  - 1点のみで推定
  - 例：母平均μを標本平均で推定
- **区間推定**
  - 幅を持たせて推定
  - 「平均μは確率0.95でAとBの間にある」などと言う

平均の区間推定

σ:既知とする.

$N(\mu, \sigma^2)$ からの大きさ n の無作為標本を,$x_1, x_2, \ldots, x_n$ とする.

正規分布をしている母集団の平均 μ の区間推定

- 手順1：標本平均を求める

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \qquad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

※標本平均 $\bar{x}$ が平均 μ, 分散 σ^2/n の正規分布にしたがう

- 手順2：変数の基準化を行う

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}, \quad Z \sim N(0, 1)$$

平均の区間推定

- 手順3： 標準正規分布の95%点からzの信頼区間を構成

$$-1.96 < \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < 1.96$$

- 手順4：

手順３で得た信頼区間をμについて書き換える

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

<u>μの95%信頼区間</u>

例

- 18 歳女性の身長は, 分散 $\sigma^2 = 5.0^2$ である正規分布に従う
  20 人のデータ：

$$153.921, 160.855, 158.577, 153.409, 155.558, 159.569, 157.461,$$
$$156.976, 155.497, 154.748, 159.564, 155.267, 161.852, 156.816,$$
$$149.118, 156.534, 160.164, 156.634, 158.198, 157.045$$

- 20 人のデータの平均：$\bar{x} = 156.888$.

- 母平均 μ の 95% 信頼区間：

$$154.697 = \bar{x} - 1.96\frac{5}{\sqrt{20}} < \mu < \bar{x} + 1.96\frac{5}{\sqrt{20}} = 159.079$$

標本の大きさと区間幅

- 標本数が大きくなるにつれ、信頼区間の幅が狭くなる。すなわち、標本数が大きくなると、高い精度でμを推定できる.

標準偏差を5とした場合の標本数と信頼区間の関係

標本数	5	25	100	500
信頼幅	8.77	3.92	2.00	0.88

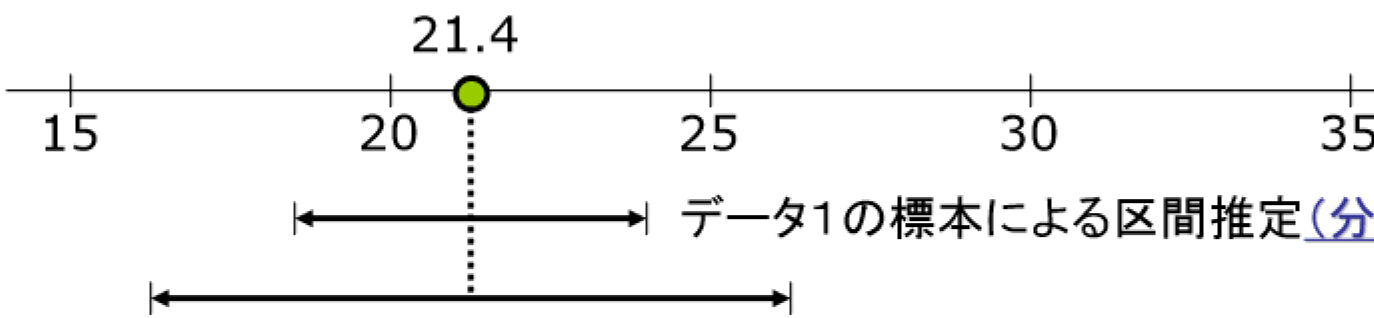
※ 信頼幅 $= 2 \times 1.96 \boxed{\dfrac{\sigma}{\sqrt{n}}}$

標本数が大きくなるにつれ，小さくなる

区間推定

区間推定のイメージ

どの程度の正確さで母数を推定しているかを
その区間が示している

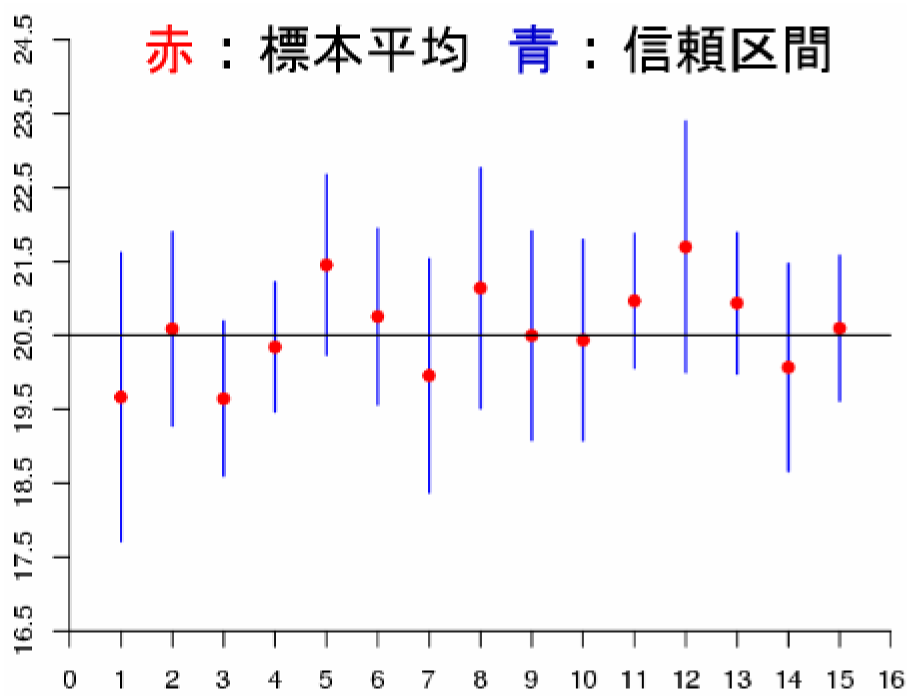


データ1の標本による区間推定(分散が小さい)

データ2の標本による区間推定(分散が大きい)

区間推定に関するシミュレーション

- 平均20.5，分散4の正規分布にしたがう乱数を15個生成し，その標本平均と95%信頼区間を求めた
- 繰り返し回数：15回



分散が未知の場合

これまでは、母分散が既知として信頼区間を求めた

- 一般には母分散は未知

母分散 σ^2 を s^2 で推定し，次の統計量を構成

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

統計量 t は自由度 $n-1$ の t 分布にしたがう

※正規分布ではない

分散が未知の場合の区間推定

自由度 m の t 分布の両側 5%点を $t_{0.05}(m)$ とすると，
t 統計量の 95%信頼区間は

$$-t_{0.05}(m) < \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n}}\right)} < t_{0.05}(m)$$

※ $m = n - 1$

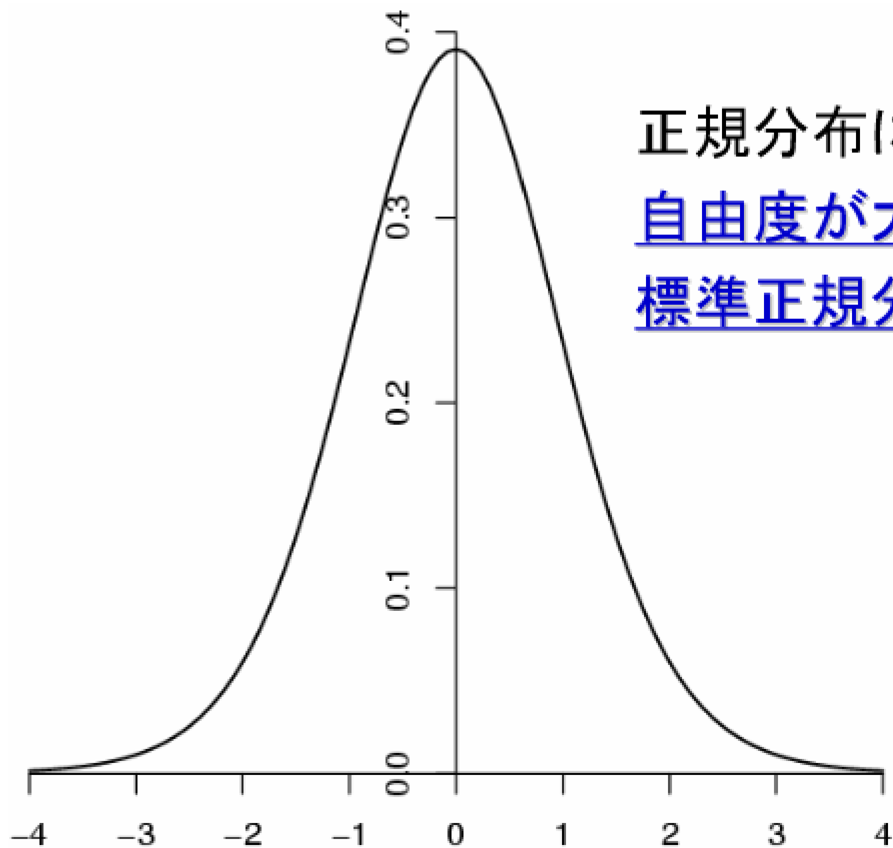
これを μ について解いて

$$\bar{x} - t_{0.05}(m)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.05}(m)\frac{s}{\sqrt{n}}$$

分散が未知の場合の μ の95%信頼区間

t分布

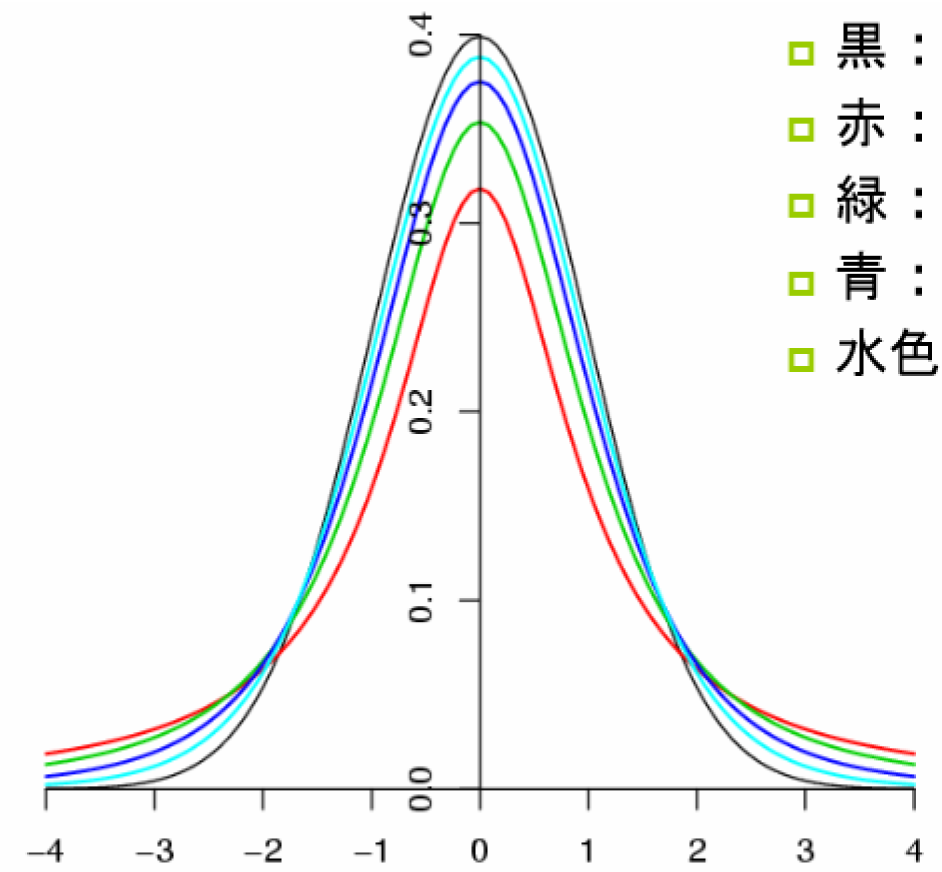
- 下の図は，自由度12の*t* 分布を描いたものである



正規分布によく似ており、
自由度が大きくなるにつれて
標準正規分布に近づいていく

t分布の自由度

- 自由度1，2，4，9のt 分布と標準正規分布



- 黒：標準正規分布
- 赤：自由度 1
- 緑：自由度 2
- 青：自由度 4
- 水色：自由度 9

例

- 目的
  - 花びらの長さの母平均μの95%信頼区間を求める
- データ
  - ひあふぎあやめを20本採取し，花びらの長さを測定

    1.4, 1.5, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5

    1.5, 1.6, 1.4, 1.1, 1.2, 1.5, 1.3, 1.4, 1.7, 1.5 (cm)
- μの95%信頼区間

$$\bar{x} - t_{0.05}(m)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.05}(m)\frac{s}{\sqrt{n}}$$

例

- 標本平均

$$\bar{x} = \frac{1.4 + 1.5 + \cdots + 1.5}{20} = 1.44$$

- 標本分散・標準偏差

$$s^2 = \frac{(1.4 - 1.44)^2 + (1.5 - 1.44)^2 + \cdots + (1.5 - 1.44)^2}{20 - 1}$$

$$= 0.0215$$

- t 統計量の両側5%点 （付表や統計パッケージから得られる）

$$t_{0.05}(19) = 2.09$$

例

- 信頼区間を求める.

$$\bar{x} - t_{0.05}(19)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.05}(19)\frac{s}{\sqrt{n}}$$

$$\bar{x} = 1.44, \quad s = 0.147, \quad t_{0.05}(19) = 2.09$$

より

$$1.44 - 2.09\frac{0.147}{\sqrt{20}} < \mu < 1.44 + 2.09\frac{0.147}{\sqrt{20}}$$

よって、μ の95%信頼区間は

$$\boxed{1.37 < \mu < 1.51}$$

5.母集団分布がベルヌーイ分布の場合

ベルヌーイ分布

母集団の中で、ある性質をもつものの比率 p の推定を考える。

ある性質を持てば 1, もたなければ 0 の値をとるような確率変数を X と定義すると，母集団における X の確率分布は**ベルヌーイ分布**と呼ばれる**確率分布**になる.

$$\Pr(X = 1) = p,\ \Pr(X = 0) = 1 - p$$

- ベルヌーイ分布に従う**確率変数**は, **2 値** (例えば, 0 または,1) を取る.
- 期待値：$\mathrm{E}[X] = p$, 分散：$\mathrm{Var}[X] = p(1 - p)$.

標本平均の分布

母集団

ベルヌーイ分布

ある都市の全世帯における

ある意見に対する賛成率 p

無作為抽出

標本

n 世帯抽出

$(x_1, x_2, \ldots, x_n)$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

p を推定.

標本平均の分布

ベルヌーイ分布に従う大きさ n の無作為標本の総和を $Y = n\bar{X}$ とすると

$$\Pr(Y = k) = {}_nC_k \times p^k(1-p)^{n-k}$$

$$(k = 0, 1, 2, \cdots, n)$$

となる．このような確率分布を**２項分布**とよび，$\mathrm{B}(n, p)$ と表す．ここで ${}_nC_k$ は n 個の中から k 個取り出す組み合わせの数で

$${}_nC_k = \frac{n!}{k!(n-k)!} \quad (\text{ただし}\, n! = 1 \times 2 \times \cdots \times n, \quad 0! = 1)$$

によって計算できる（$n!$ は n の階乗と読む）．

例 （n が大きいとき、なぜ近似が必要か？）

歪みのないサイコロを 500 回振るとき，1 の目が出る回数が 80 回以上 100 回以下となる確率はいくらぐらいでしょう？

確率変数 X を 1 の目が出る回数とすると，求めたい確率は，

$$\Pr(80 \leq X \leq 100)$$

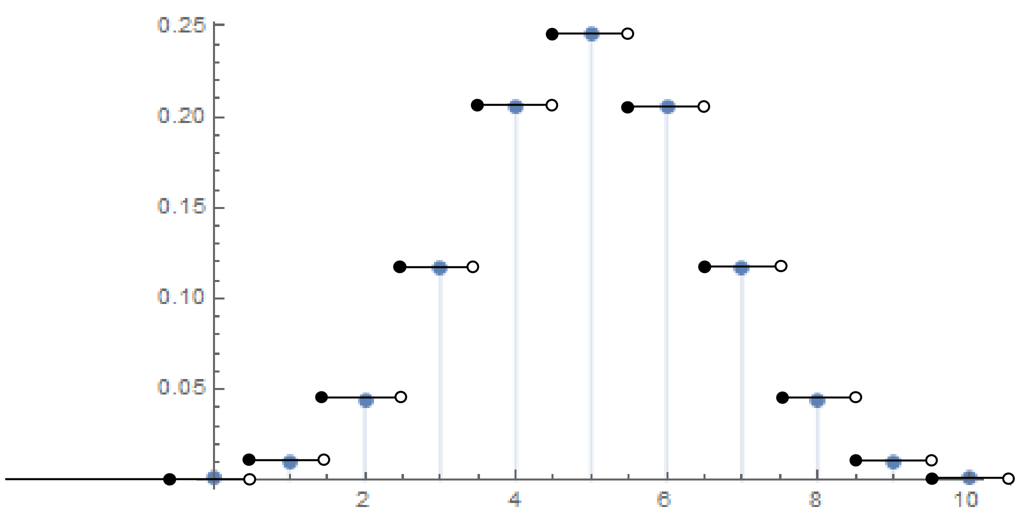
$X \sim B\left(500, \frac{1}{6}\right)$ であるから，

$$\begin{aligned}\Pr(80 \leq X \leq 100) &= {}_{500}\mathrm{C}_{80}\left(\frac{1}{6}\right)^{80}\left(\frac{5}{6}\right)^{420} + {}_{500}\mathrm{C}_{81}\left(\frac{1}{6}\right)^{81}\left(\frac{5}{6}\right)^{419} \\ &\quad + \cdots + {}_{500}\mathrm{C}_{100}\left(\frac{1}{6}\right)^{100}\left(\frac{5}{6}\right)^{400}\end{aligned}$$

${}_{500}\mathrm{C}_{80}, {}_{500}\mathrm{C}_{81}, \ldots, {}_{500}\mathrm{C}_{100}$ の計算は大変である!

実は、n が大きい場合、二項分布と正規分布は関係がある

$$
\begin{cases}
\vdots \\
\vdots \\
0 \quad \text{その他}
\end{cases}
$$



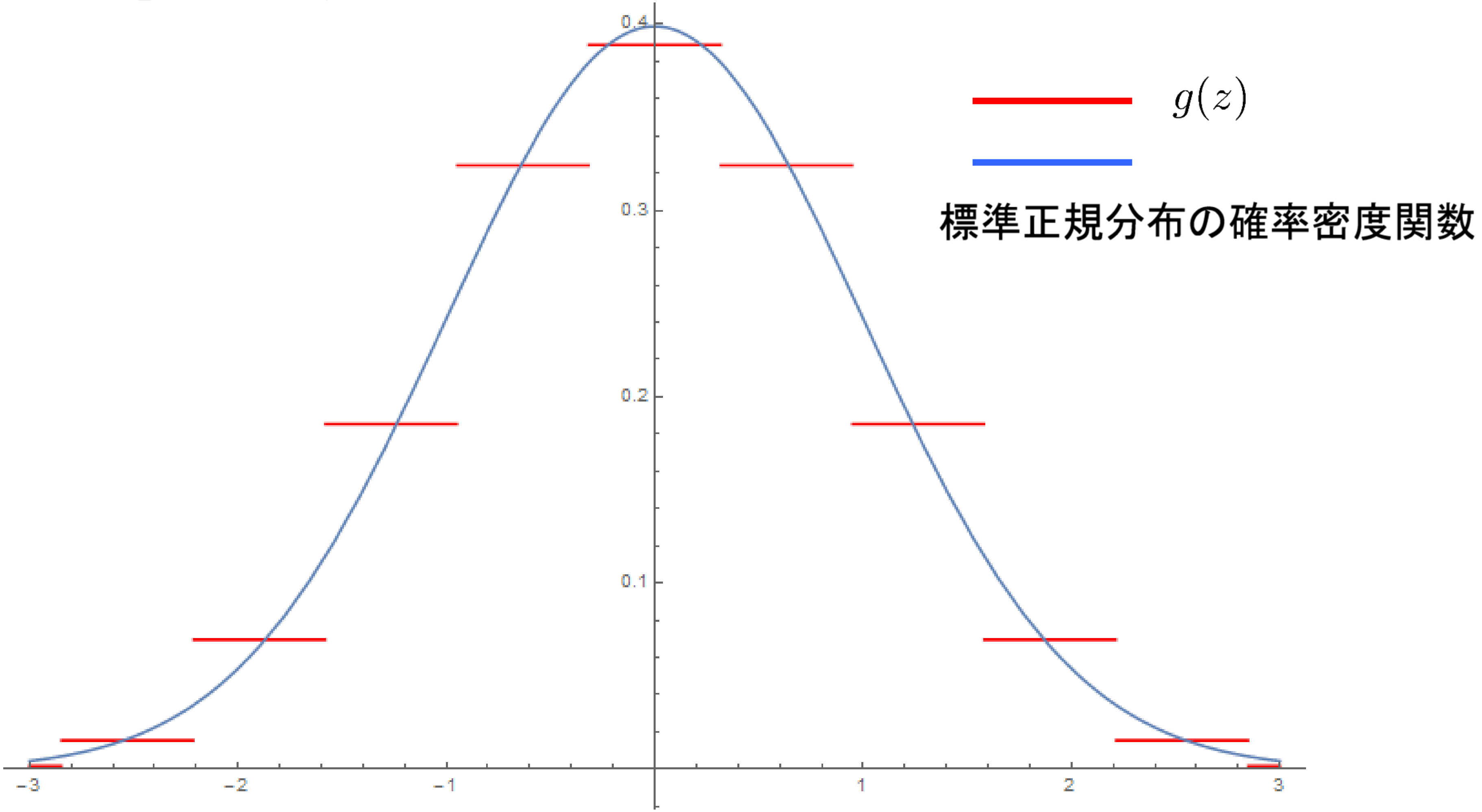
ZとN(0,1)の関係について考えよう

X の確率密度関数を用いて，$Z = (X - np)/\sqrt{npq}$ の確率密度関数を求めると以下のようになる．

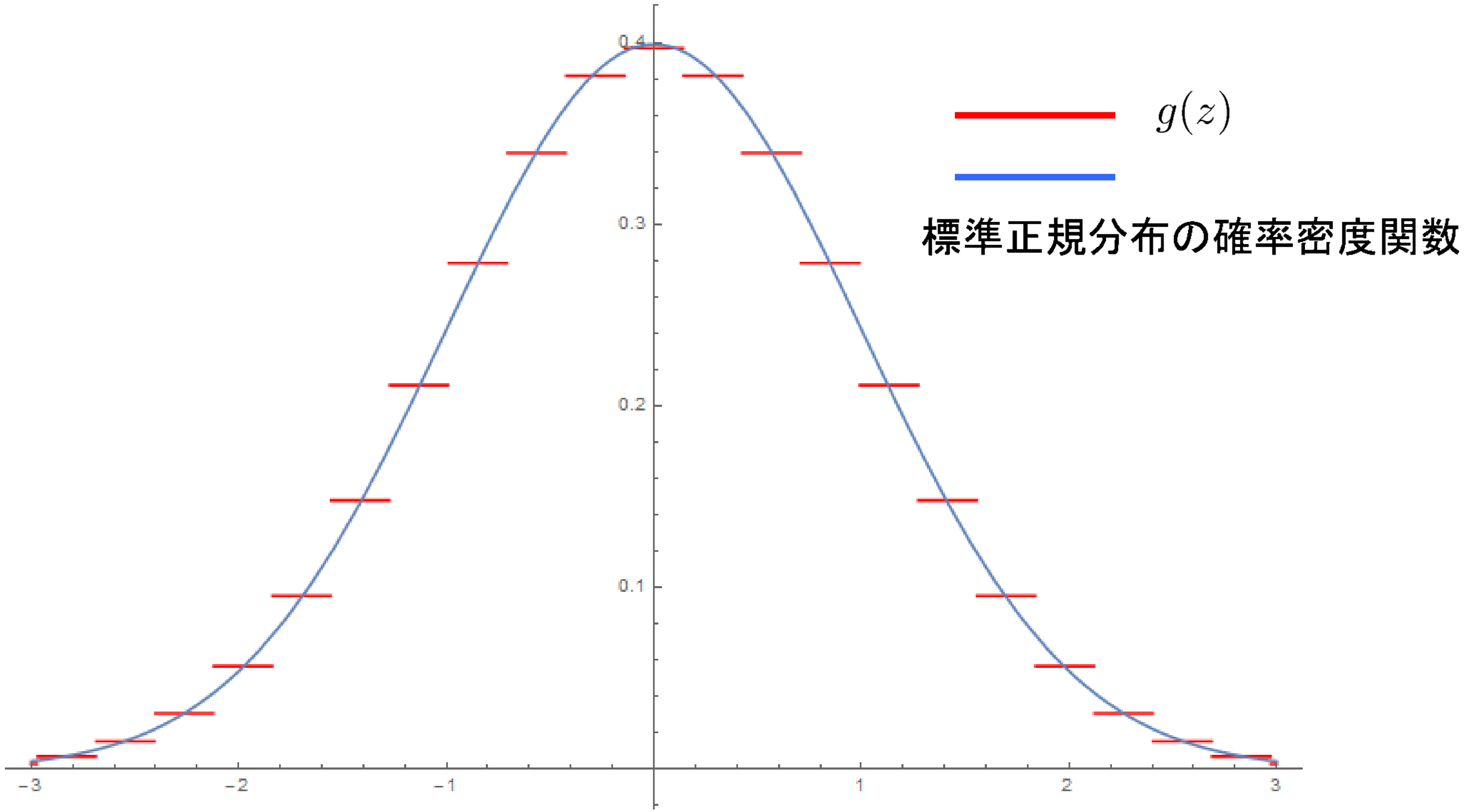
$$
g(z) = \begin{cases}
\sqrt{npq} \times {}_n\mathrm{C}_0\, p^0 (1-p)^{n-0}, & \dfrac{0 - \frac{1}{2} - nq}{\sqrt{npq}} \leq z < \dfrac{0 + \frac{1}{2} - np}{\sqrt{npq}} \\
\sqrt{npq} \times {}_n\mathrm{C}_1\, p^1 (1-p)^{n-1}, & \dfrac{1 - \frac{1}{2} - np}{\sqrt{npq}} \leq z < \dfrac{1 + \frac{1}{2} - np}{\sqrt{npq}} \\
\vdots & \\
\sqrt{npq} \times {}_n\mathrm{C}_k\, p^k (1-p)^{n-k}, & \dfrac{k - \frac{1}{2} - np}{\sqrt{npq}} \leq z < \dfrac{k + \frac{1}{2} - np}{\sqrt{npq}} \\
\vdots & \\
\sqrt{npq} \times {}_n\mathrm{C}_n\, p^n (1-p)^{n-n}, & \dfrac{n - \frac{1}{2} - np}{\sqrt{npq}} \leq z < \dfrac{n + \frac{1}{2} - np}{\sqrt{npq}} \\
0 & \text{その他}
\end{cases}
$$

数値例（赤と青は似ているか？？）

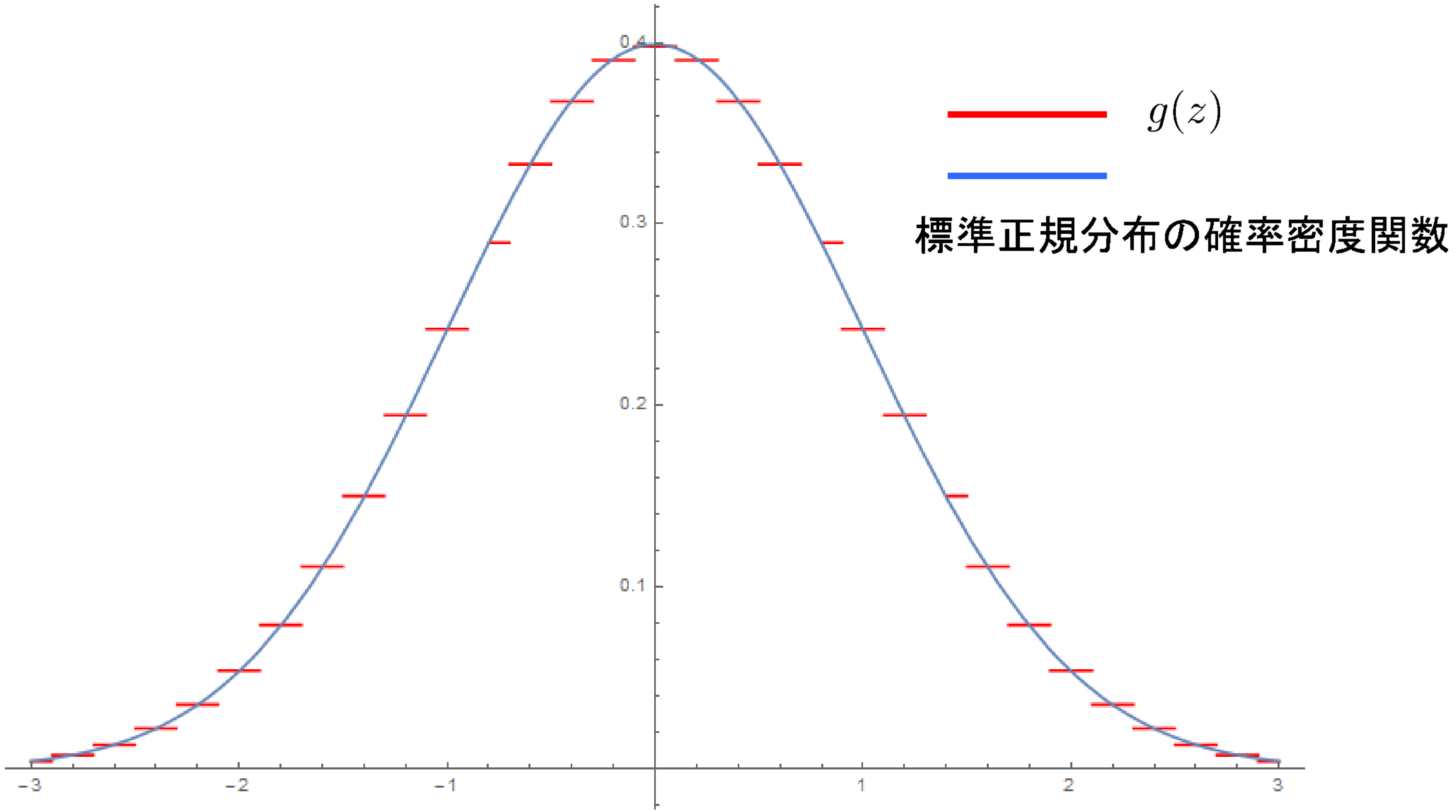
$p = 0.5, n = 10$



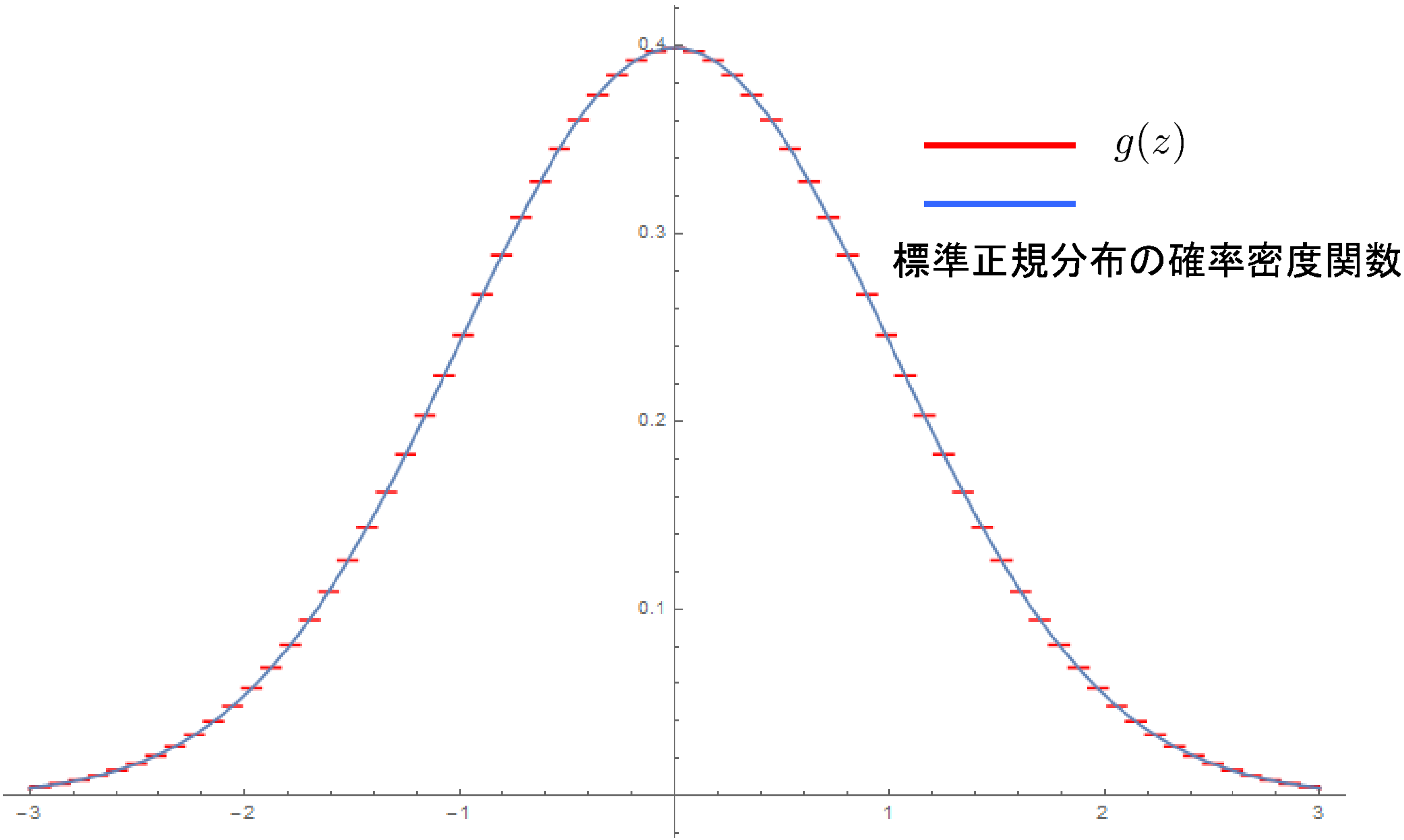
$$p = 0.5, n = 50$$



$$p = 0.5, n = 100$$



$p = 0.5, n = 500$



n が大きい場合、二項分布≒正規分布

n が非常に大きい場合は, 基準化した二項分布 $g(x)$ は標準正規分布で近似できる.

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

「X は近似的に正規分布に従う」という意味は２項分布のおよその確率が正規分布の確率で求められることを表している．このことを２項分布の**正規近似**という.

このことは, 中心極限定理によって正当化される!

中心極限定理(母集団分布が2項分布)

定理 [標本平均の極限分布]

$p \in (0,1)$ を固定し,$n \to \infty$ の下で,

$$Y_n = \frac{\sum_{i=1}^{n}(X_i - p)}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}}$$

の分布関数は $\mathcal{N}(0,1)$ の分布関数へ収束する.

実用上は, 分母の $p(1-p)$ を推定した分布を与えておく必要がある.

分散の推定量

実用上, $p(1-p)$ を推定する必要がある．推定量として,

$$\bar{X}(1-\bar{X})$$

を考える．

実用化された統計量 $Z_n = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}}$ の正規近似の妥当性を示すには中心極限定理だけではなく以下のような定理を使う．

定理 [標本平均の極限分布]

補題 [推定量の一致性]

定理 [スラツキーの定理]

定理 [連続写像定理]

$$\Rightarrow \quad Z_n = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} \approx \mathcal{N}(0,1)$$

分散の推定量の一致性

補題 [推定量の一致性]

任意の $\varepsilon > 0$ に対して,

$$\Pr(|\bar{X}(1-\bar{X}) - p(1-p)| > \varepsilon) \to 0 \ as \ n \to \infty$$

この事実を、

$$\bar{X}(1-\bar{X}) \xrightarrow{P} p(1-p)$$

と表し、$\bar{X}(1-\bar{X})$ は $p(1-p)$ へ確率収束するという.

スラッキーの定理

$\{X_n\}_{n=1}^{\infty}, X$ は, 確率変数, $F_X(x)$ を X の分布関数とする．このとき, $F_X(x)$ の任意の連続点において

$$\lim_{n\to\infty} \Pr(X_n \leq x) = F_X(x)$$

であることを X_n が X へ分布収束するといい,

$$X_n \xrightarrow{\mathcal{D}} X$$

と表す.

定理 [スラッキーの定理]

$\{X_n\}_{n=1}^{\infty}, X, \{Y_n\}_{n=1}^{\infty}$ を確率変数, $c \in \mathbb{R}, c \neq 0$ とする．このとき,

$$X_n \xrightarrow{\mathcal{D}} X, Y_n \xrightarrow{P} c \Rightarrow \frac{X_n}{Y_n} \xrightarrow{\mathcal{D}} \frac{X}{c}$$

連続写像定理

定理 [連続写像定理]

確率変数の列を $\{X_n\}_{n=1}^{\infty}$ とし, $\Pr(X_n \notin I) \to 0$, $c \in I$ とする.
I で連続な関数を $g(x)$ とする. このとき,

$$X_n \xrightarrow{P} c \Rightarrow g(X_n) \xrightarrow{P} g(c)$$

である.

実用化された標本平均の漸近分布

$\bar{X}(1-\bar{X}) \xrightarrow{P} p(1-p)$ （$\because$ 補題 [推定量の一致性] ）と連続写像定理より,

$$\sqrt{\frac{\bar{X}(1-\bar{X})}{p(1-p)}} \xrightarrow{P} 1 \ as \ n \to \infty \quad (*)$$

定理 [標本平均の極限分布] より,

$$\frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \ as \ n \to \infty \quad (**)$$

いま,

$$Z_n = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}} \Big/ \sqrt{\frac{\bar{X}(1-\bar{X})}{p(1-p)}}$$

であるから,$(*)$, $(**)$ とスラツキーの定理より次の結果を得る.

実用化された標本平均の漸近分布

定理 [標本平均の漸近分布]

$p \in (0,1)$ を固定し,$n \to \infty$ の下で,

$$Z_n = \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{\bar{X}(1-\bar{X})}}$$

の分布関数は $\mathcal{N}(0,1)$ の分布関数へ収束する.

Note:分散に未知パラメイタ p を含んでいない。

比率の近似信頼区間

ベルヌーイ分布をしている母集団の平均p の区間推定

- 手順1：標本平均を求める

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

- 手順2：変数の基準化を行う

$$z = \frac{\sqrt{n}(\bar{x} - p)}{\sqrt{\bar{x}(1 - \bar{x})}}$$

比率の近似信頼区間

- 手順3： 標準正規分布の 95% 点から z の信頼区間を構成

$$-1.96 < \frac{\sqrt{n}(\bar{x}-p)}{\sqrt{\bar{x}(1-\bar{x})}} < 1.96$$

- 手順4：

手順３で得た信頼区間を p について書き換える

$$\bar{x} - 1.96\frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}} < p < \bar{x} + 1.96\frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}}$$

μ の95%信頼区間

中心極限定理の一般化について

確率変数列 $X_1, X_2, \cdots, X_n$ の和 $X_1 + X_2 + \cdots + X_n$ の $n \to \infty$ とした場合の分布について言及しているのが中心極限定理と言える.

- 確率変数列は独立で, それらの母集団分布は同一で平均と分散が存在を仮定

リンドバーグ＝レヴィの中心極限定理

- 母集団分布の同一性の仮定を緩めた結果

リャプノフ型中心極限定理

- 確率変数列の独立性を緩めた結果

マルチンゲール中心極限定理

中心極限定理【リャプノフ型】

以下の定理は, 中心極限定理の一般化である.
母集団分布の同一性の仮定を緩めている.

定理 [リャプノフの中心極限定理]

$X_1, X_2, \cdots, X_n$ を独立な確率変数列とする. いま, 各 i に対して, $\mathrm{E}[X_i] = \mu_i$, $\mathrm{E}[(X_i - \mu_i)^2] = \sigma_i \neq 0$, $\mathrm{E}[|X_i - \mu_i|^k] = \beta_i$(ある $k \in \mathbb{N}$, $k \geq 3$) が存在するとする. さらに,

$$B_n = \left(\sum_{i=1}^{n} \beta_i\right)^{1/k}, \quad C_n = \left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}$$

に対して, $\lim_{n\to\infty}(B_n/C_n) = 0$ ならば,

$$Y_n = \frac{\sum_{i=1}^{n}(X_i - \mu_i)}{C_n}$$

の分布関数は $\mathcal{N}(0,1)$ の分布関数へ収束する.

第 100 回全国算数･数学教育研究東京大会

2018年8月3日　麹町中学にて

御清聴ありがとうございました。