医療統計学 vol.6

木村　朗

「二項分布」から正規分布、標準正規分布につなげる

二項分布

- サイコロを5回続けて投げたとき、1の目がでる回数をXとする
- 投げる回数　5としたが＞　一般にn回とし、nを大きくしていったら・・・
- 確率変数Xの確率の分布、B（n、1/6）は・・・

＞＞投げる回数nを増やしていくと・・・

次第に・・・左右対称の山状になっていく・・・

これは一般の二項分布B（n、p）についても成立する.

二項分布B（n、p）

- 平均E（X）＝np

- 分散V（X）＝npq

  但し　q＝1-p　　　（p＋q＝１だからね）

$$E(aX+bY)=aE(X)+bE(Y)$$
$$E(XY)=E(X)E(Y)$$

正規分布とは

- 日本の成年男子の身長のばらつき･･･全体を考える

データの数をどんどん増やす･･･
それに伴い、階級の幅を限りなく小さくする･･･
この時のグラフの形は･･･

ほぼ左右対称な山状になる
この時の方程式は･･･

$y=(1/\sqrt{(2\pi\sigma^2)}) \times e^{-A}$

ここで$A=((X-m)^2)/2\sigma^2$

$(-\infty < X < \infty)$

π　＝3．1415926･･･

e　＝2．71828･･･

mは実数の値　　mean･･･の意味

σは正の実数をとる

この式であらわされる分布は、平均m、標準偏差σの正規分布という

$N(m,\sigma^2)$

★確率変数が連続の場合は、Σ記号を積分記号分散で置き換えて次のように表す。

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$Var(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

離散値による確率変数による分布は、期待値を確率の**総和**で求め、分布の面積が得られたが、
連続値による場合、総積を求める・・・**積分すること**によって面積が求められる。

確率変数の言葉で正規分布を表す

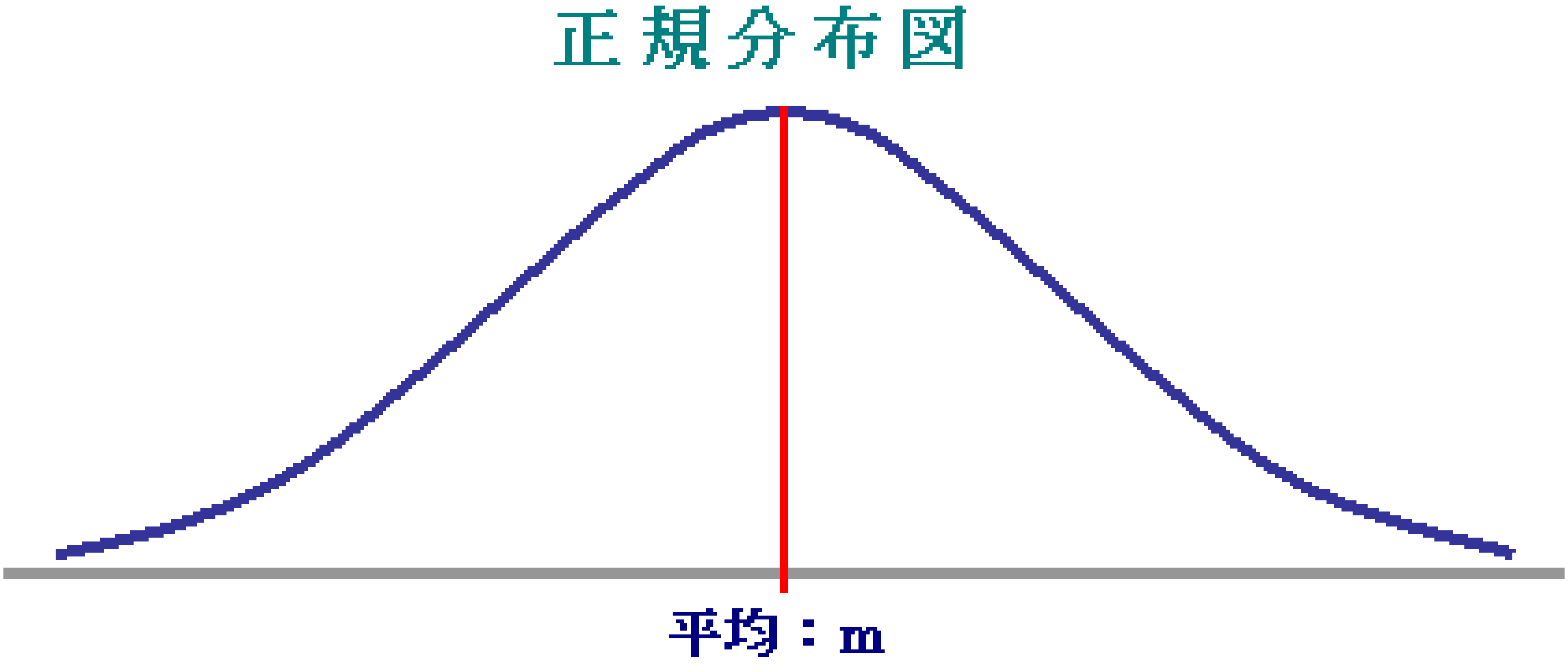
- 確率変数Xは平均m、標準偏差σの正規分布に従うといい

N（m、σ^2）という記号で表す

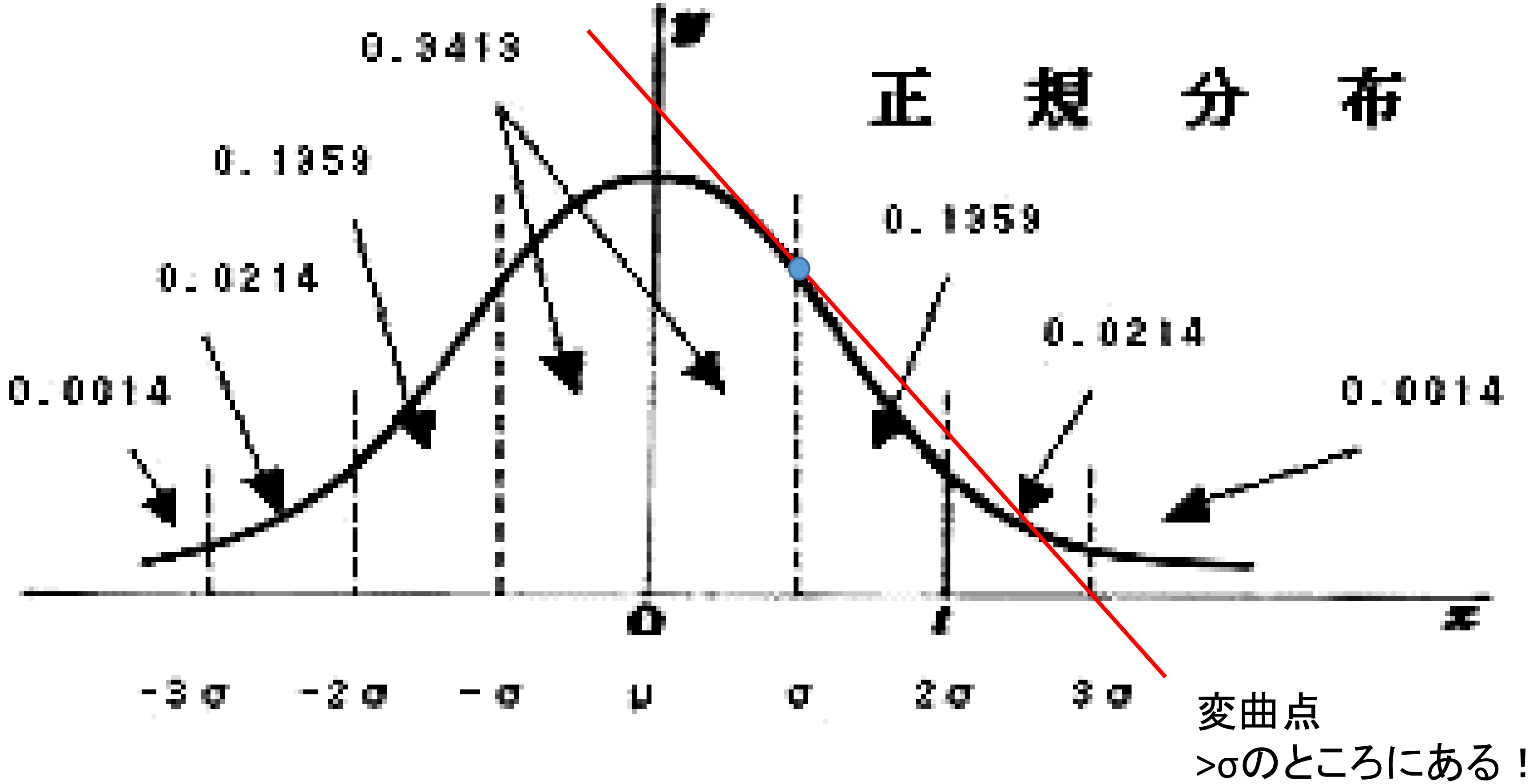
Normal　distribution
σ^2は標準偏差の２乗・・・分散

この分布は **身長・毎年の雨量・標準的なテスト・製品の工作誤差・測定誤差などの誤差**にもよく近似できる。

実用性が高い確率分布の一つ。


正規分布図
平均：m

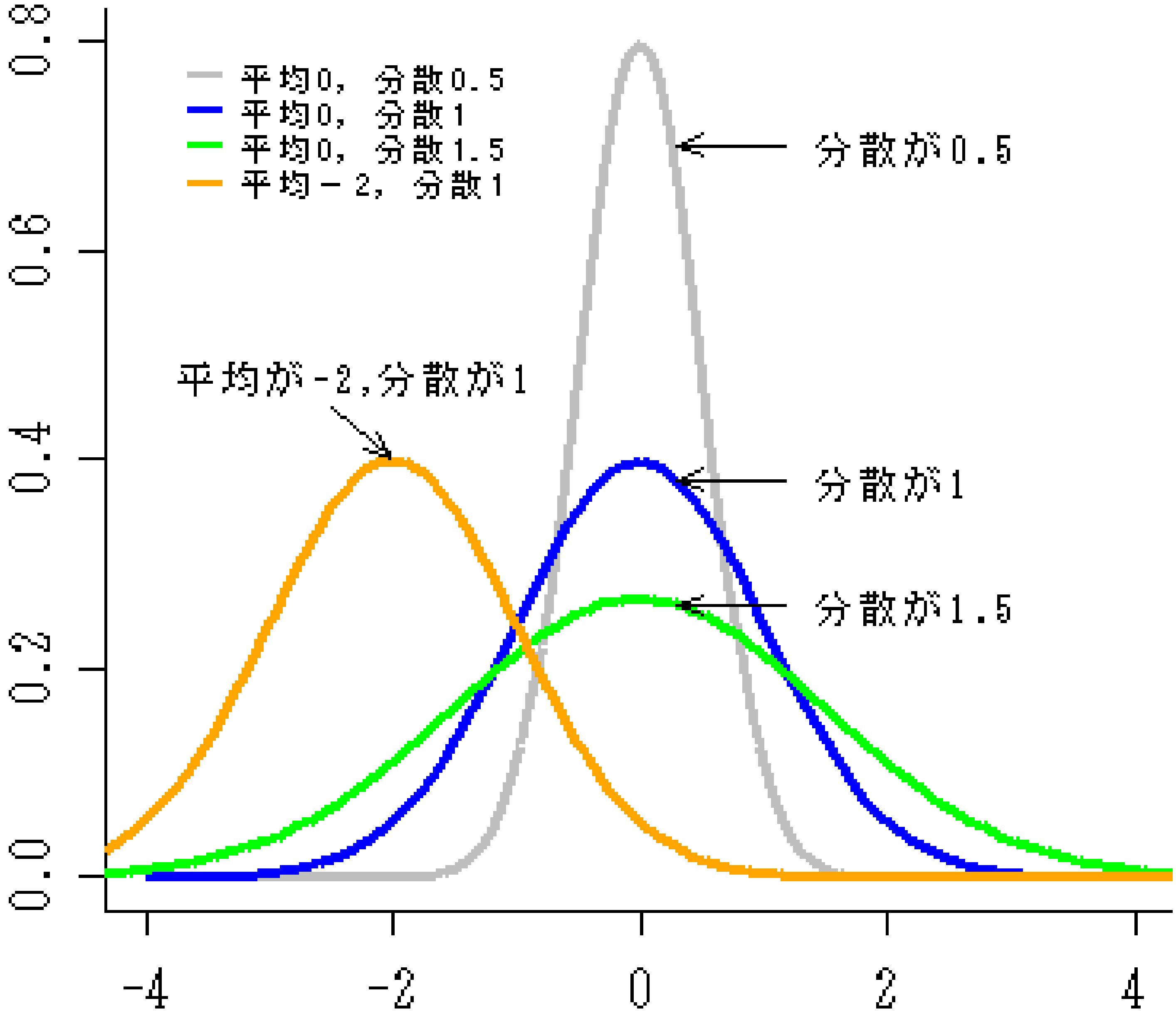
榛名山のすそ野の広がり方は、正規分布そのもの。トップが尖れば完璧なんだが・・・



正規分布は　平均と標準偏差を決めれば、分布の形が決まる。

特徴：平均mを中心として、左右対称、
サイコロの目のような飛び飛びの確率変数ではなく、もっと細かい・・・**連続的な確率変数をとる連続的確率変数**による、**連続的な確率分布**である。

全面積は**1，　±1SDに68.26%　2SDに95.44%　3SDに99.73%**



σが異なると、高さが変わる　mが変われば頂点の位置が変わる

標準正規分布とは

- 平均m　標準偏差σ^2のN分布をさらに一ひねり

特別なケース

N(0, 1)について考える･･･

今T＝X-m/σ　と置く

Tも確率変数

Tは平均m＝0、標準偏差σ＝1の正規分布N(0, 1)を**標準正規分布**という

y＝（1/√（2πσ＾2））×　e＾-A

ここでA＝（（X－m）＾2）/2σ＾2

（－∞＜X＜∞）

T=X-m/σと置くと

y＝（1/√（2π））×e＾－B

ここでB＝X^2/2

確率変数Tがc以下となる
確率P（T＜＝c）をI（c）とおくと
I（c）＝P（T＜＝c）
境界線は面積に影響を与えない
P（T＜＝c）＝P（T＜c）

π　＝3．1415926･･･

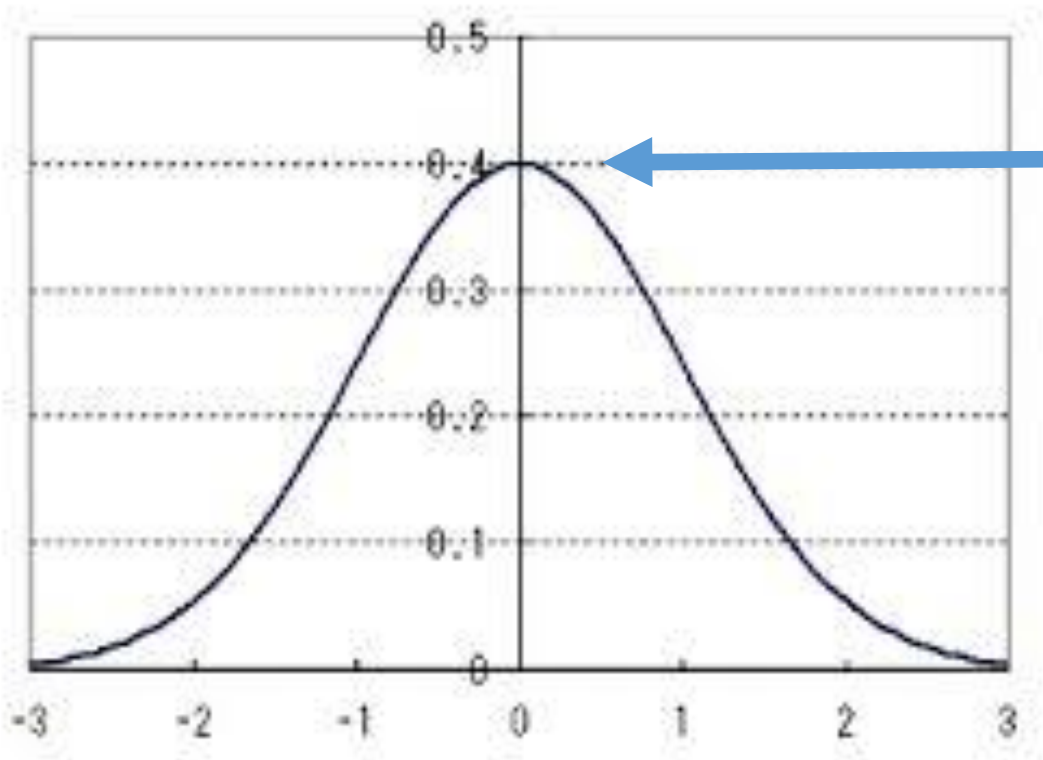
e　＝2．71828･･･

mは実数の値　　mean･･･の意味

σは正の実数をとる

この式であらわされる分布は、平均m、標準偏差σの正規分布という

標準正規分布の性質



1/√2π

Near ＝0.4

確率変数Tが標準正規分布に従うとき、P（T＜＝c）＝I（c)はどうなるのか？

- I（O）＝P（T＜＝O）＝0.5
- I（-c)＝P（T＜＝-c）＝P（T＞＝c）
- P（T＞＝c）＝１－P（T＜＝c）＝１－I（c）
- I（c）＋I（-c）＝１

- P（T＜＝c）＝P（T＜c）、P（T＞＝c）＝P（T＞c）

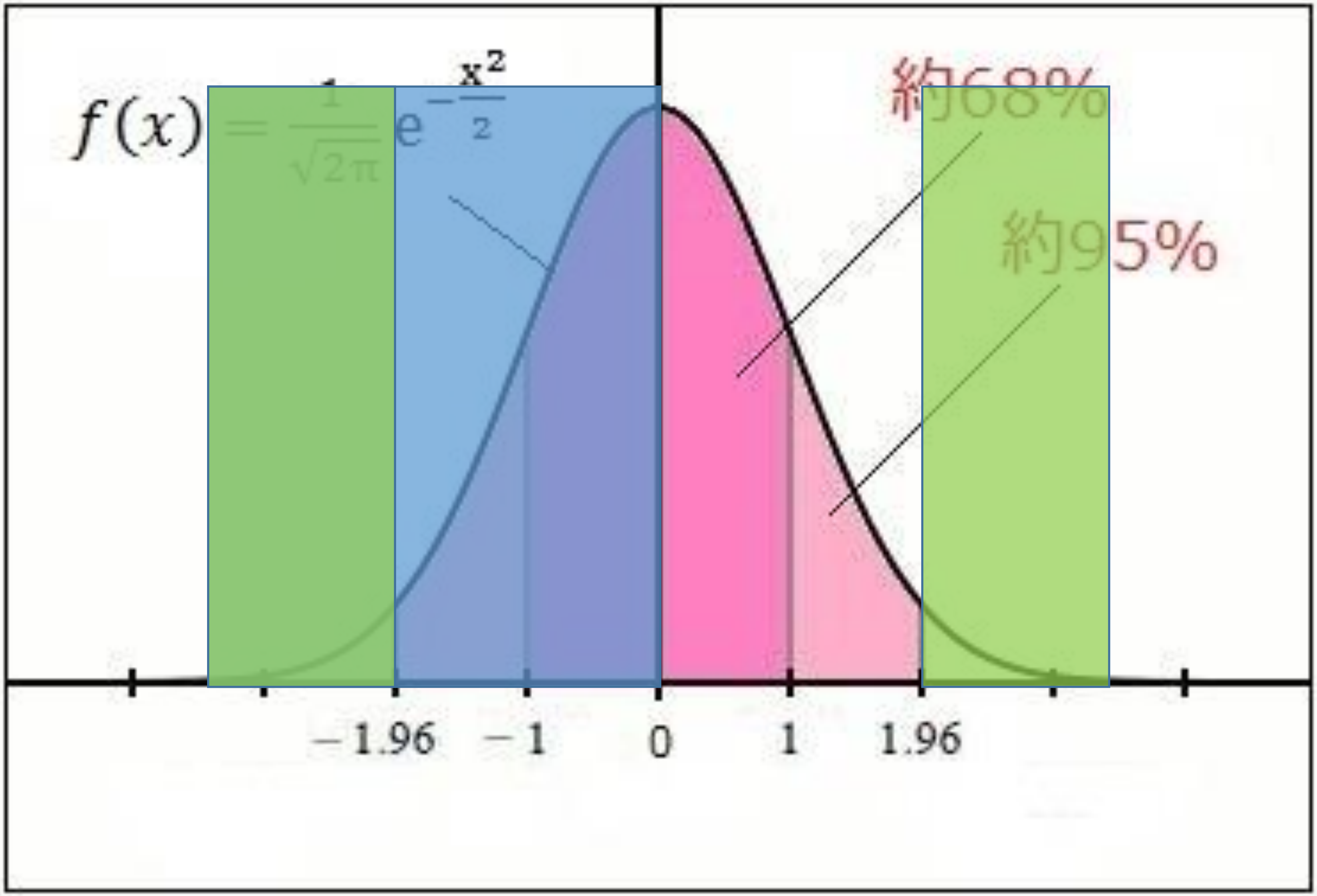
Ｃ＝1.96の時

P（T＜＝-1.96）＝I（-1.96）＝0.025

P（T＞＝1.96）＝１（-1.96）＝0.025

P（-1.96＜T＜1.96）＝I（1.96）-I（-1.96）＝0.95

・・・全面積＝0.025＋0.025＋0.95＝１



$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
約68%
約95%
−1.96
−1
0
1
1.96

標準正規分布を活用する

大学教員のAさんの身長175cm。
ある国からBさんが訪ねてくることになった。
ある国の大人の平均身長は182cm、標準偏差は8.3cm。
ある国のBさんの身長がAさんより高い確率を求める。
Bさんの大人の身長をXとする、
T＝X－m/σ＝X-182/8.3 この時、Tは標準正規分布に従う、{X＞175}という事象は
T＞175-182/8.3＝-0.843
P（X＞175）＝P（T＞-0.84）－１－P（T＜＝-0.84）＝
１－I（-0.84）＝1-0.2005＝0.80
求める確率は約0.80（80%）と計算できる。

課題

- サイコロを3回投げた時、2以下の目が出る回数をXとする、確率変数Xの分布を求めよ。