CC BY

データサイエンス概論I＆II

確率と確率分布

九州大学 数理・データサイエンス教育研究センター

なぜデータ分析に確率が？

むずかしそう

- （後で言いますが）ヒストグラムと同じような話です
  - 身長150cm台の人が全体の何人いますか？　（ヒストグラム≒統計）
  - 身長150cm台の人が全体の何%いますか？（確率）
  - ヒストグラムはデータの分布を理解するのに役に立つので，確率も役に立つ！
- （これも後で言いますが）統計的検定の基礎も確率
  - 「グループAとB，どれぐらいの確率で『違いがある』と言ってよいか？」という話
- 人工知能の基礎も実は確率
  - その背後には，「世の中の物事や人間の判断が確率的」という話が…
- （実は言いませんでしたが）平均や分散，距離，回帰分析，あたりの話にも確率の考えが潜んでいたりします

そもそも確率・確率分布とは？

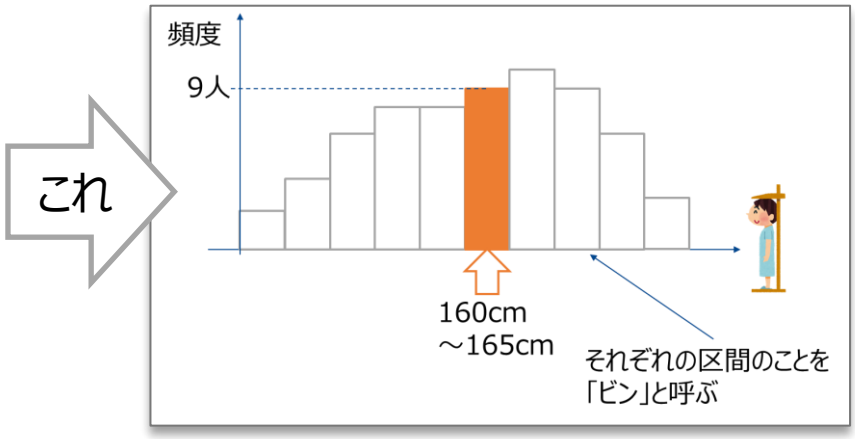
むずかしそう

確立じゃないです．確率です．

(心が折れない)
確率って何だ？

- 確率＝「出やすさ」
  - ただそれだけ．これを理解しておけば，まずは大丈夫

- 「出やすさ」なので，
  ヒストグラム(頻度)と似ている

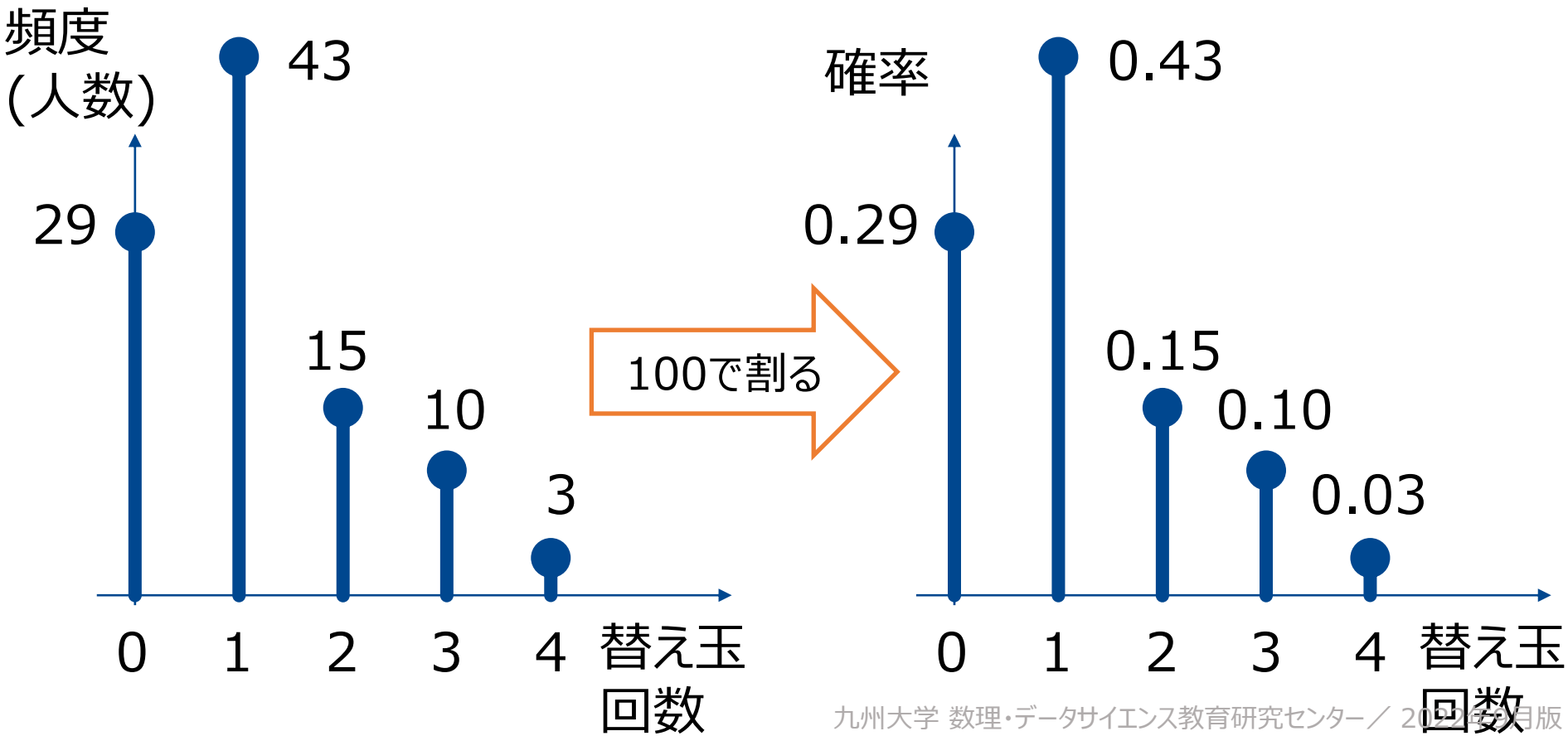


- データxが取りうる値が「有限個」と「無限個」で扱いが違う点も，ヒストグラムと似ている
  - 次々スライド以降で説明

ヒストグラムから確率分布へ：「ほとんど同じモノじゃない？」「そう，その通り！」

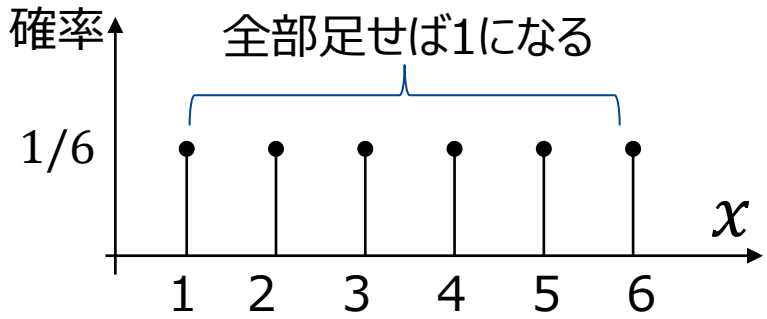
100人分ラーメン調査結果
ヒストグラム(適当に作ったデータです)

100人分ラーメン調査結果
確率分布

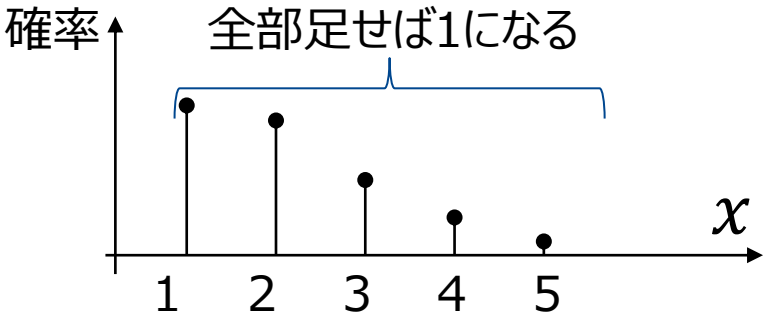


「データxの取りうる値が有限個」の場合：確率分布＝離散分布

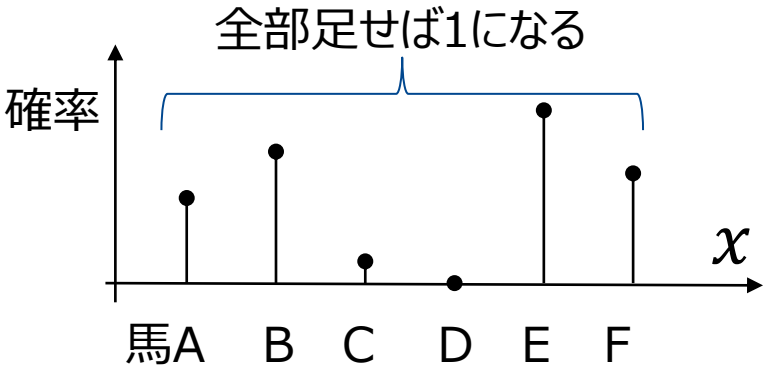
- 例１:さいころの目
  - 取りうる値＝6通り



- 例2：5段階アンケートの回答
  - 取りうる値＝5通り



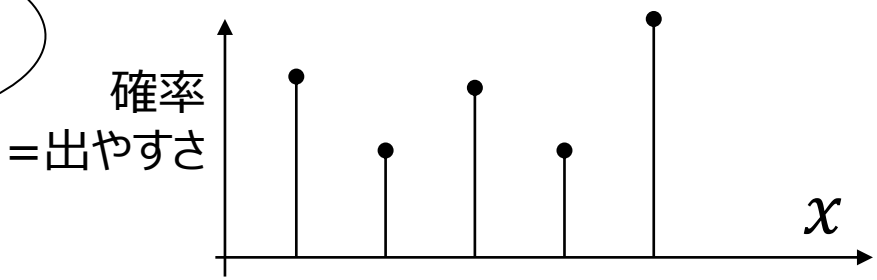
- 例３：6頭での競馬の1位
  - 取りうる値＝6通り
  - 本命 E
  - 大穴 C
  - 欠場 D



「データxの取りうる値が有限個」の場合：確率分布＝離散分布

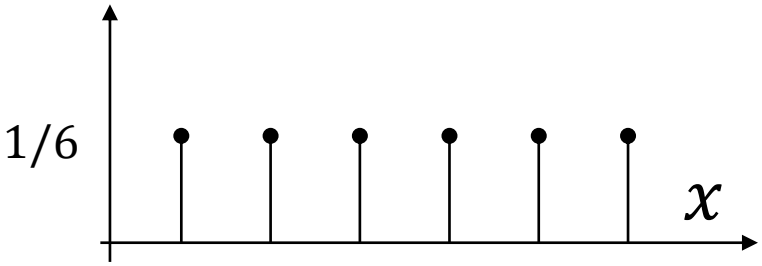
- 離散分布 $P(x)$
  - xの発生「確率」
  - ＝ xという値の出やすさ

本講義では
大文字のP

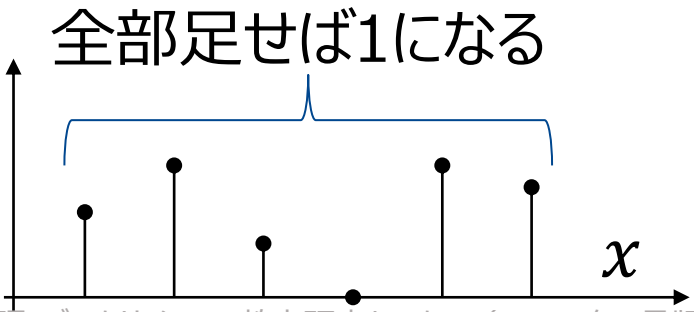


- 例：さいころの目

  - ３の目が出る確率 $P(3) = 1/6$

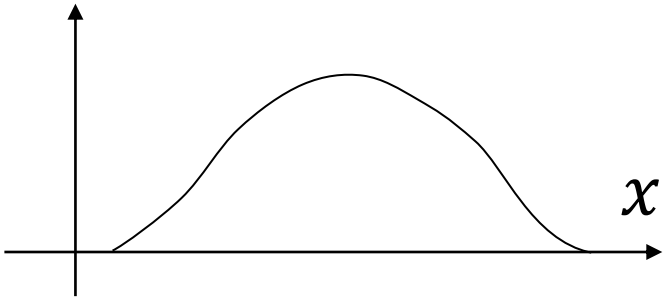


- 全部足せば1になる
  - $\sum P(x) = 1$

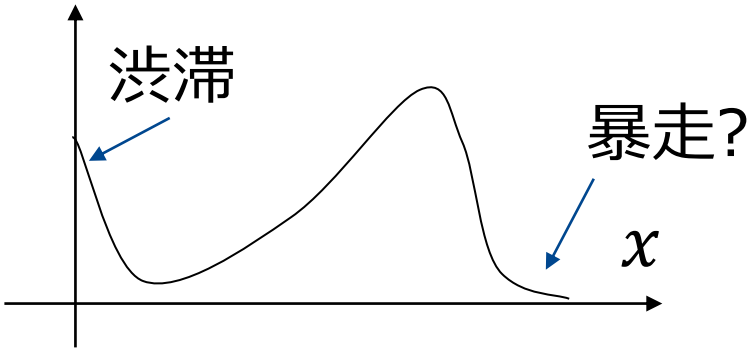


「データxの取りうる値が無限通り」の場合：連続分布

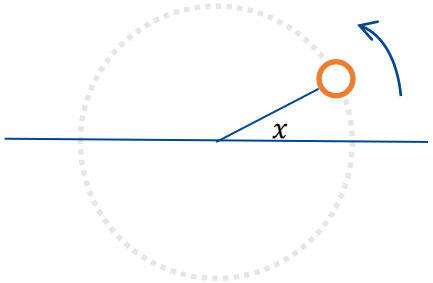
- 例１：身長x

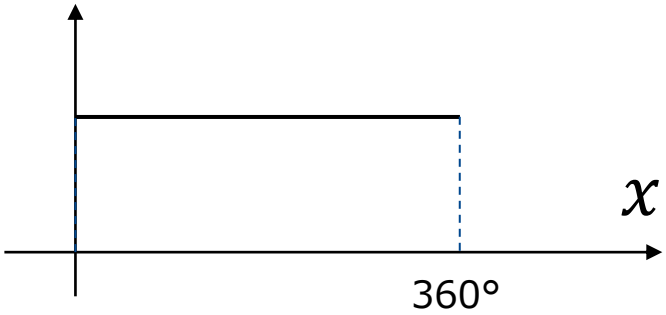


- 例2：首都高速での車速x
  - 図はイイカゲンです



- 例３：等速回転する振り子の角度x

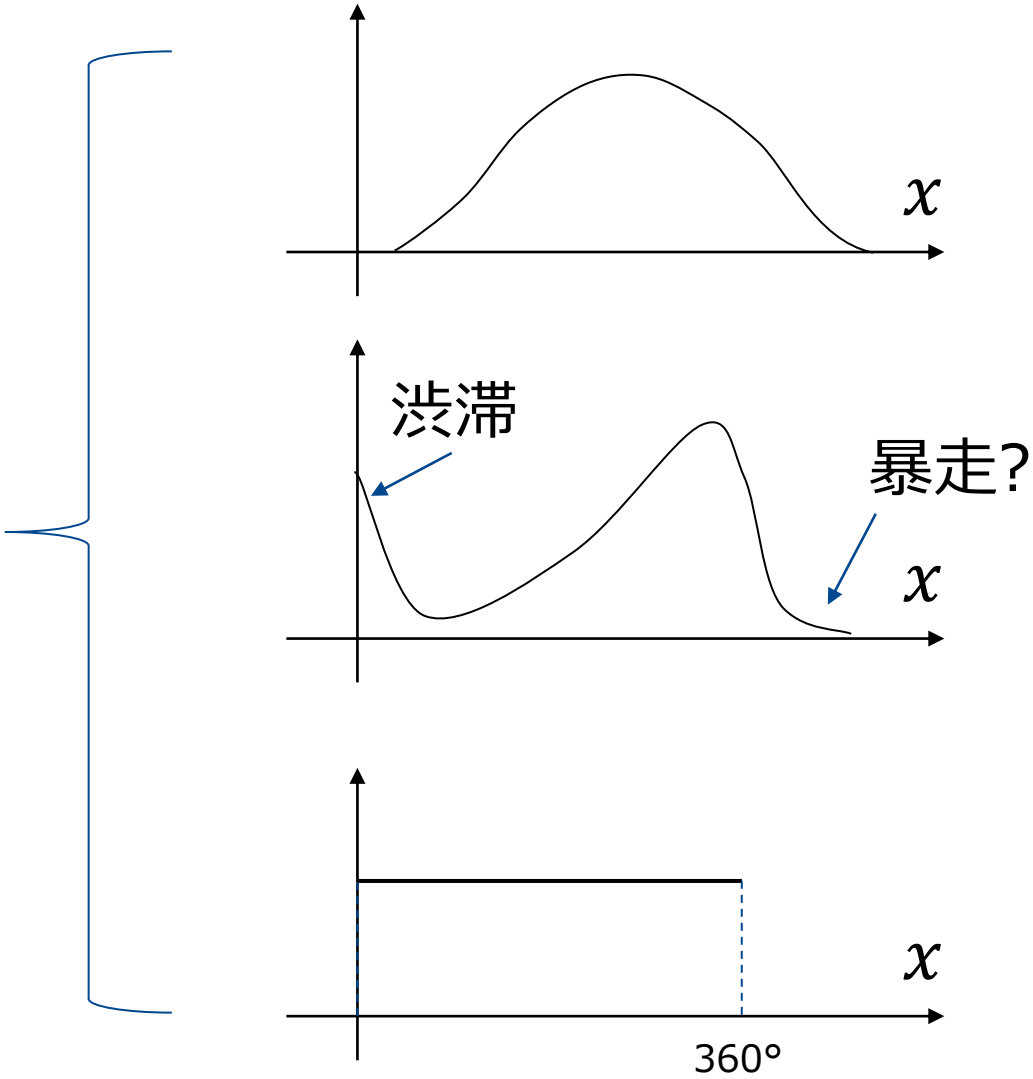




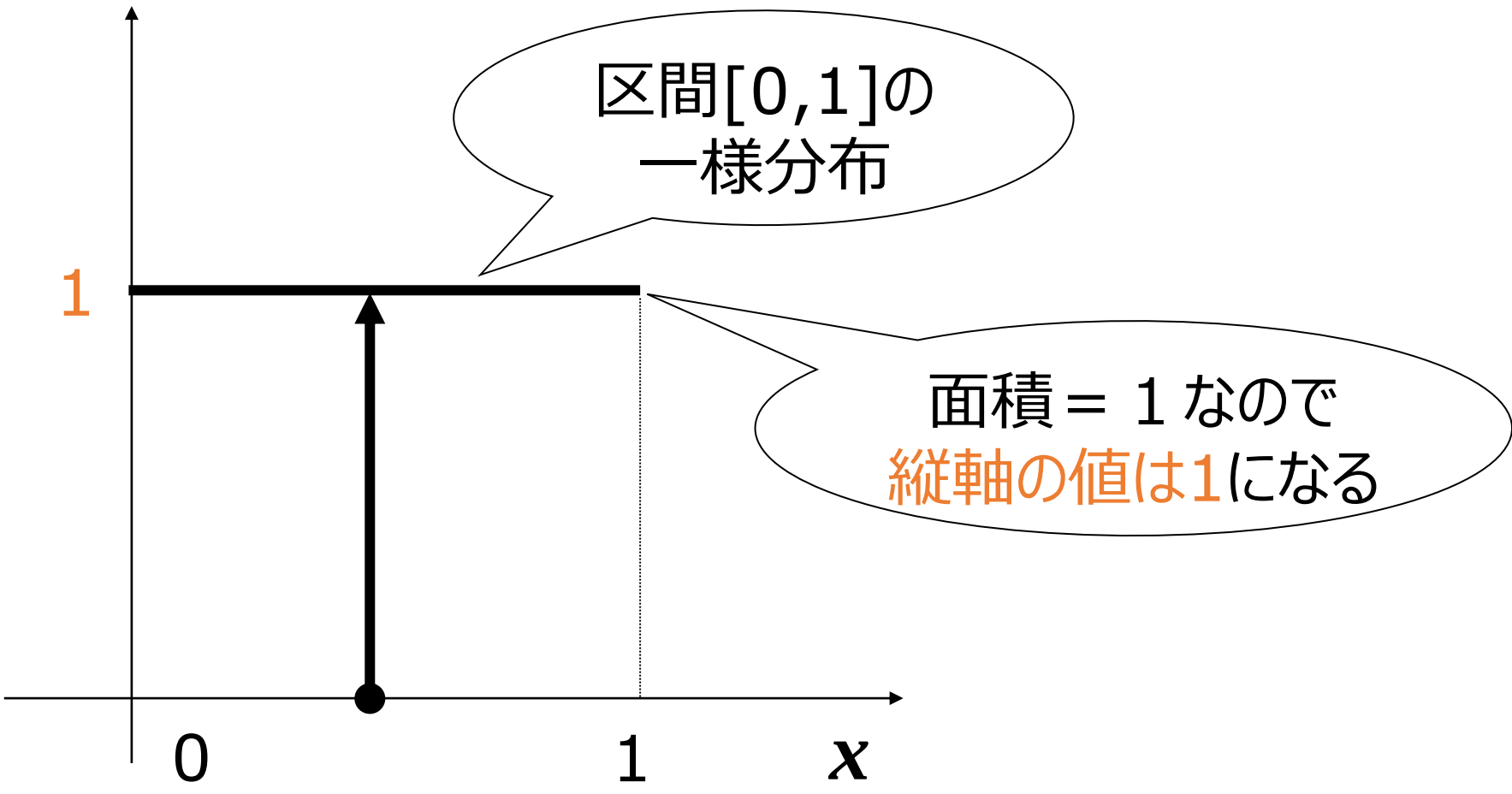
「データxの取りうる値が無限通り」の場合：連続分布

この縦軸は，
離散確率分布の
場合と同様，
「確率」でOK?

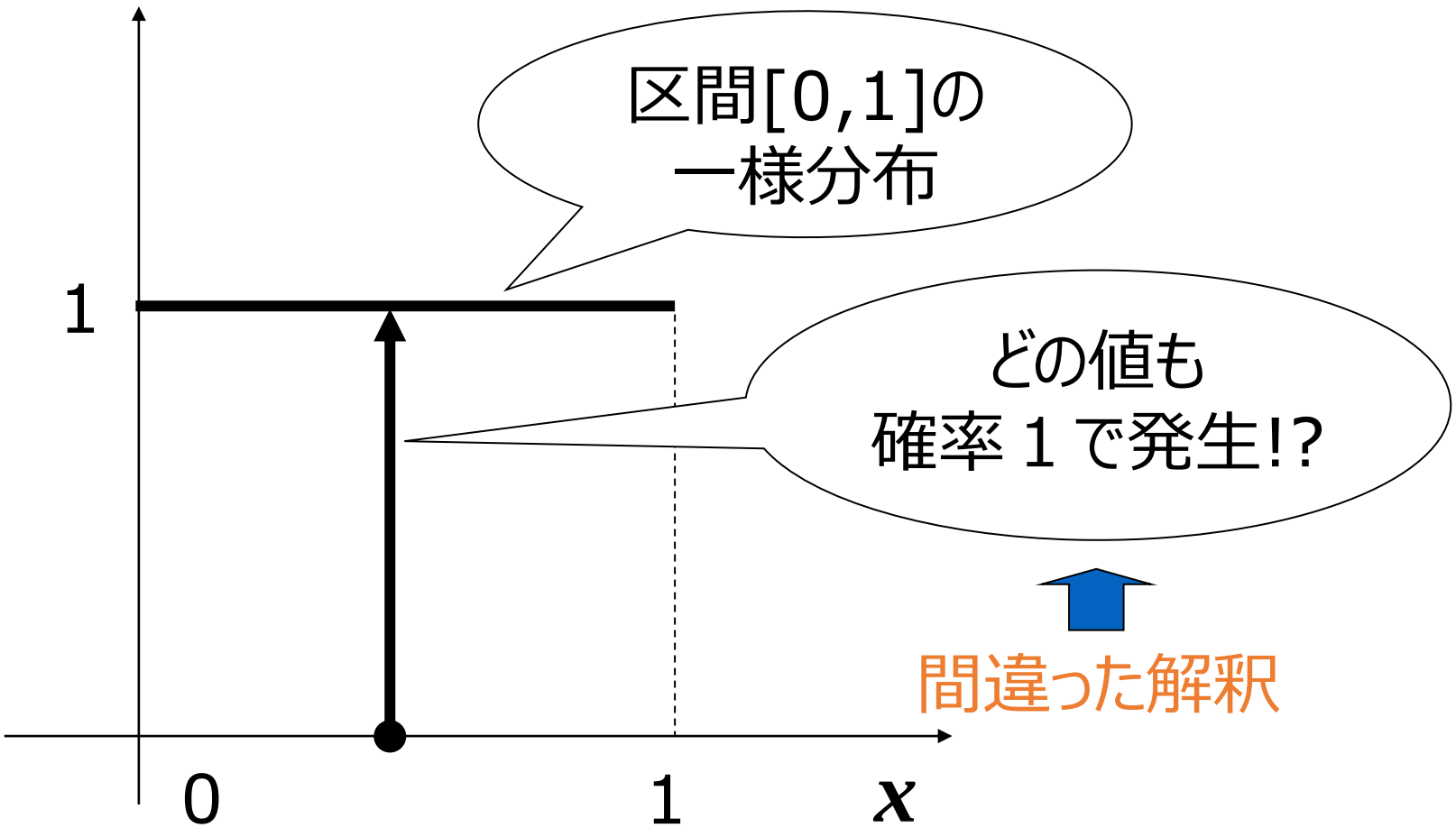
NO!



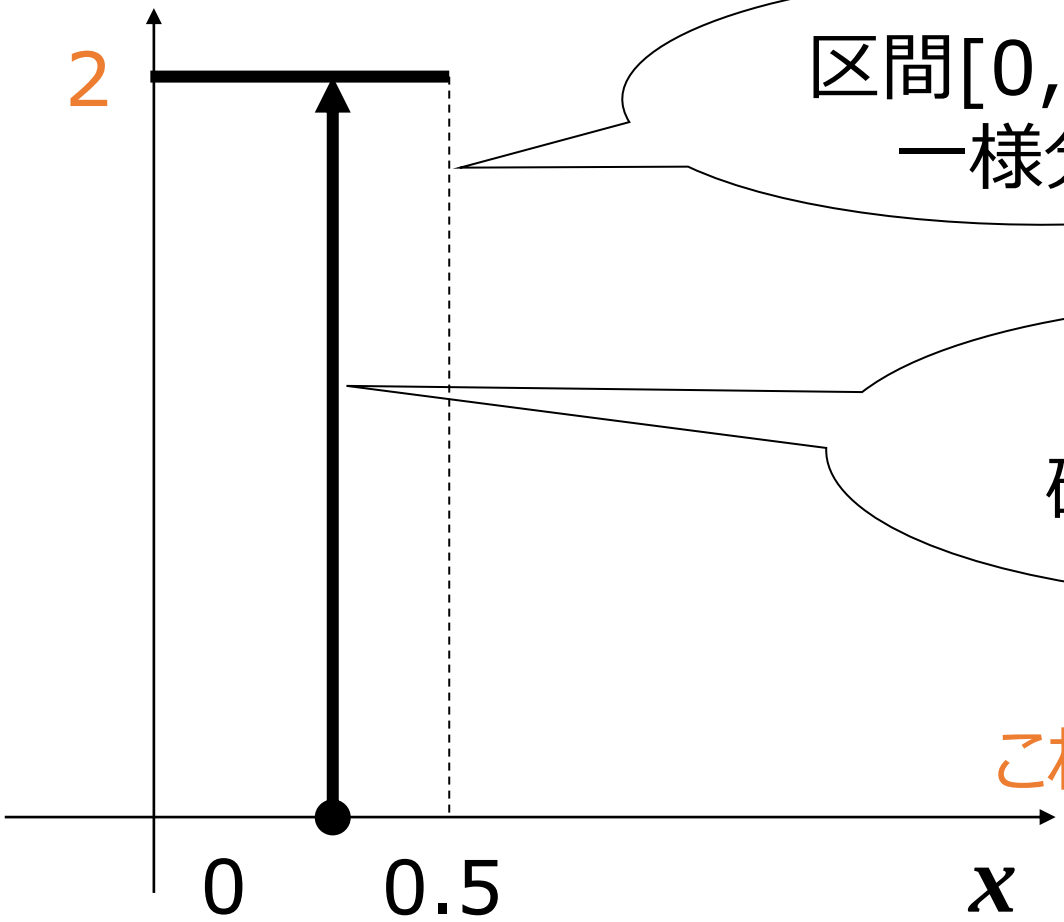
「連続分布」の縦軸が確率ではないことを理解するための例(1/4)



「連続分布」の縦軸が確率ではないことを理解するための例(2/4)



「連続分布」の縦軸が確率ではないことを理解するための例(3/4)

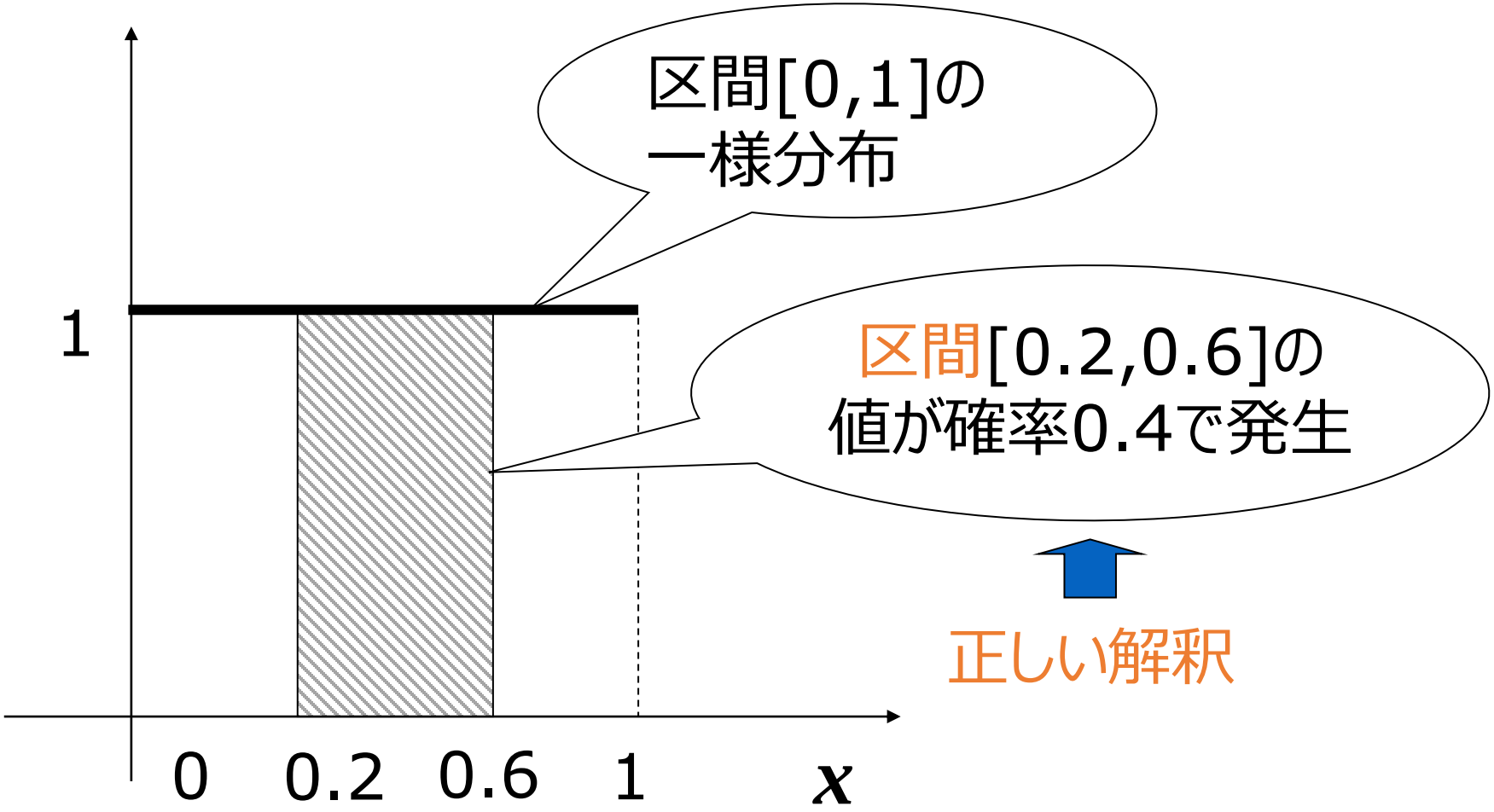


区間[0,0.5]の
一様分布

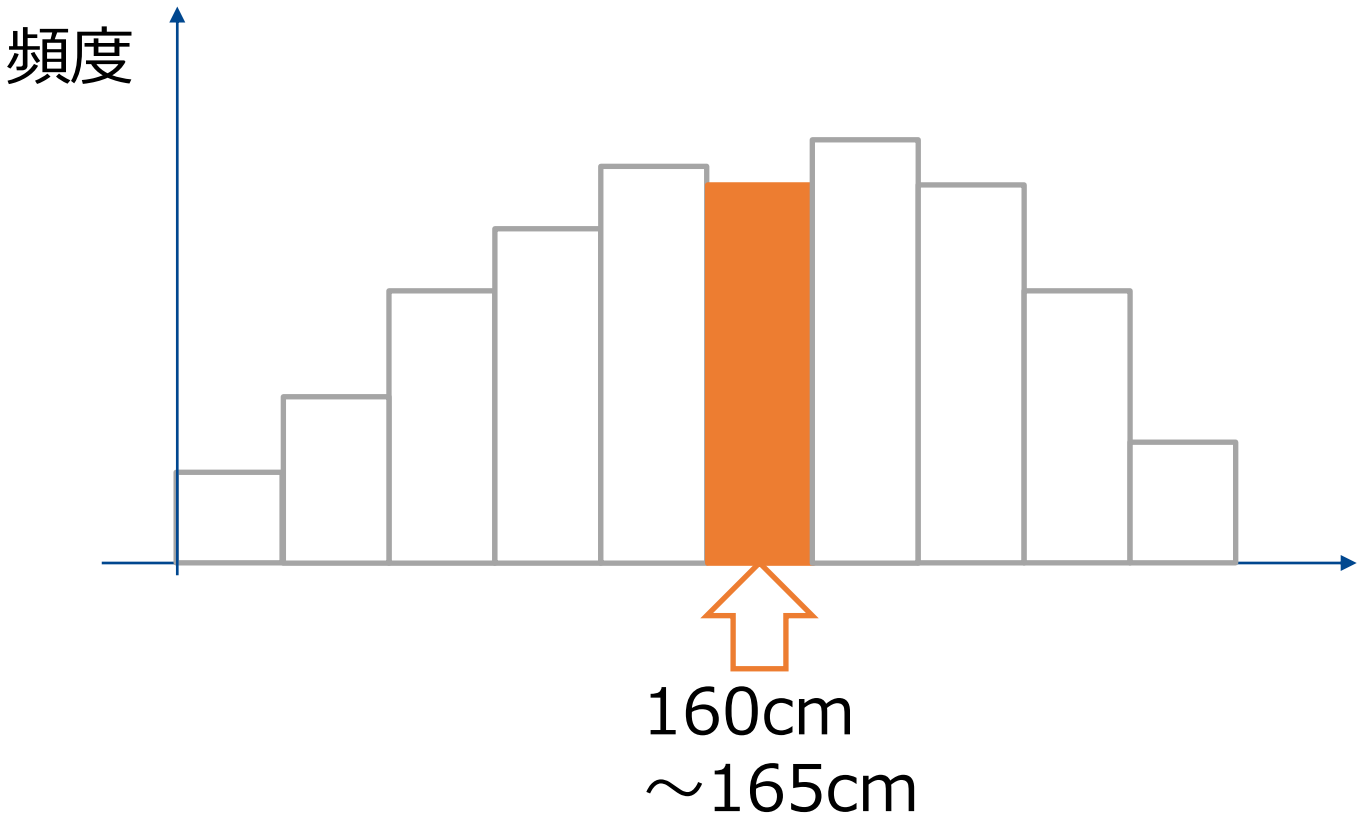
どの値も
確率2で発生!?

これも間違った解釈

「連続分布」の縦軸が確率ではないことを理解するための例(4/4)



ヒストグラムの場合と同じ！

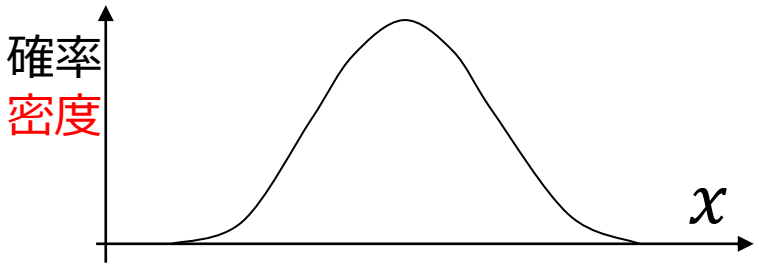


- これも「区間を考えた」から頻度が定義できたんです！

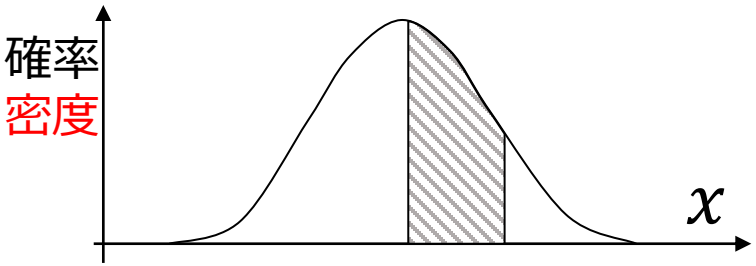
連続分布の縦軸＝「確率密度」≠「確率」

本講義では
小文字のp

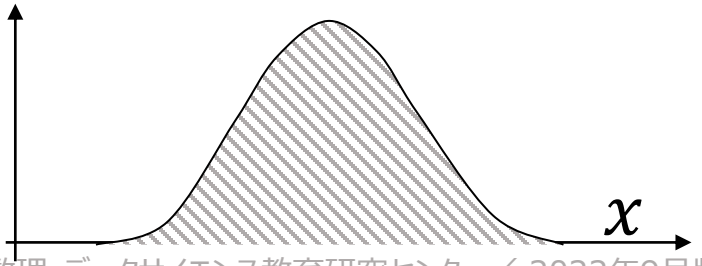
- 連続確率 $p(x)$
  - $p(x)$は確率密度



- ある区間を考えれば確率
  - 面積が確率を表す
  - 従って「値xがドンピシャで出る確率はゼロ」(わかりにくいですが…)
    - 次のスライドを参考に

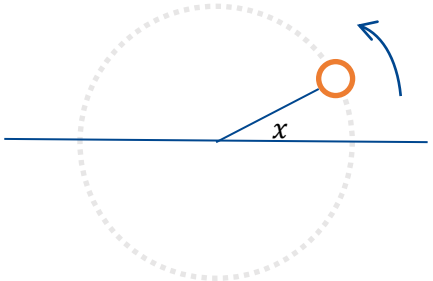


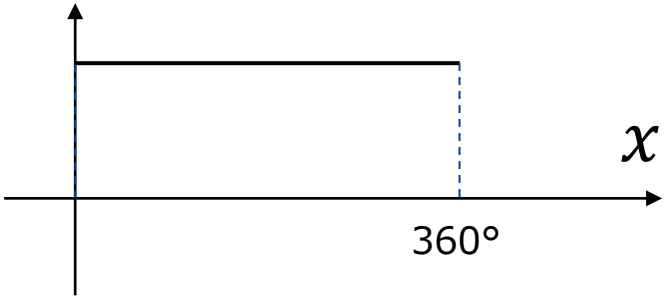
- 全部の範囲を考えれば1になる
  - $\int p(x)\,dx = 1$



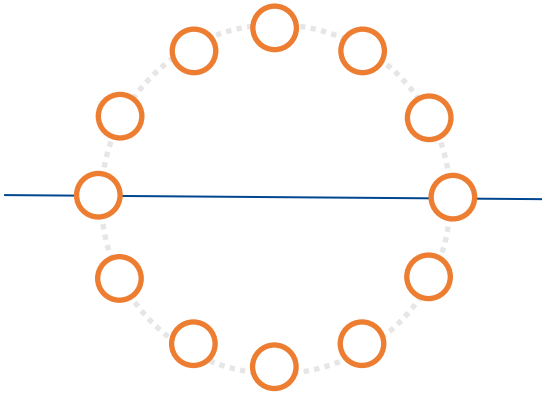
必ず起こっているのに確率ゼロというのはどういうこと？ 回転する「おもり」の角度を例に考える

- ぐるぐる回転しているので必ず$x = 90°$になる瞬間はあるはず.
- それなのに(角度は連続値なので), $x = 90°$の確率はゼロ！



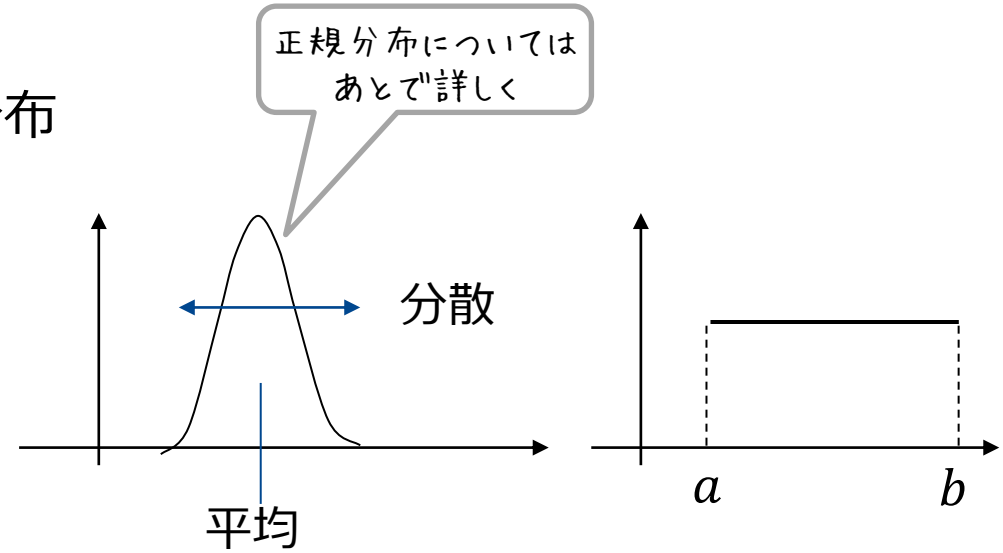


- 「全体の中そこにいる割合」と考えてはどうでしょう？
  - 右図のようにおもりの位置が全$N = 12$点のみとすれば, $x = 90°$のところにいる割合は1/12.
  - 実際には$N \to \infty$だから割合は$\to 0$
  - ちなみにxが0から90°までの範囲にいる割合とすれば, $N \to \infty$でも割合は1/4でゼロにはならない

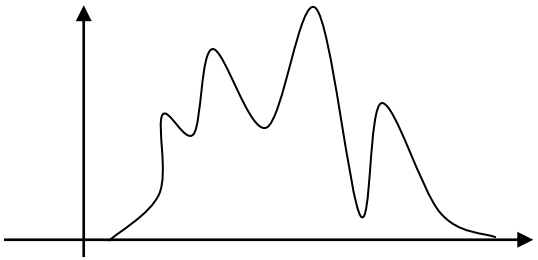


(離散・連続以外の)もう一つの分類：パラメトリックとノンパラメトリック

- パラメトリックな確率分布
  - 少数のパラメータで形状が定まる分布
  - ex. 正規分布 → 平均・分散
  - ex. 一様分布 → 分布範囲

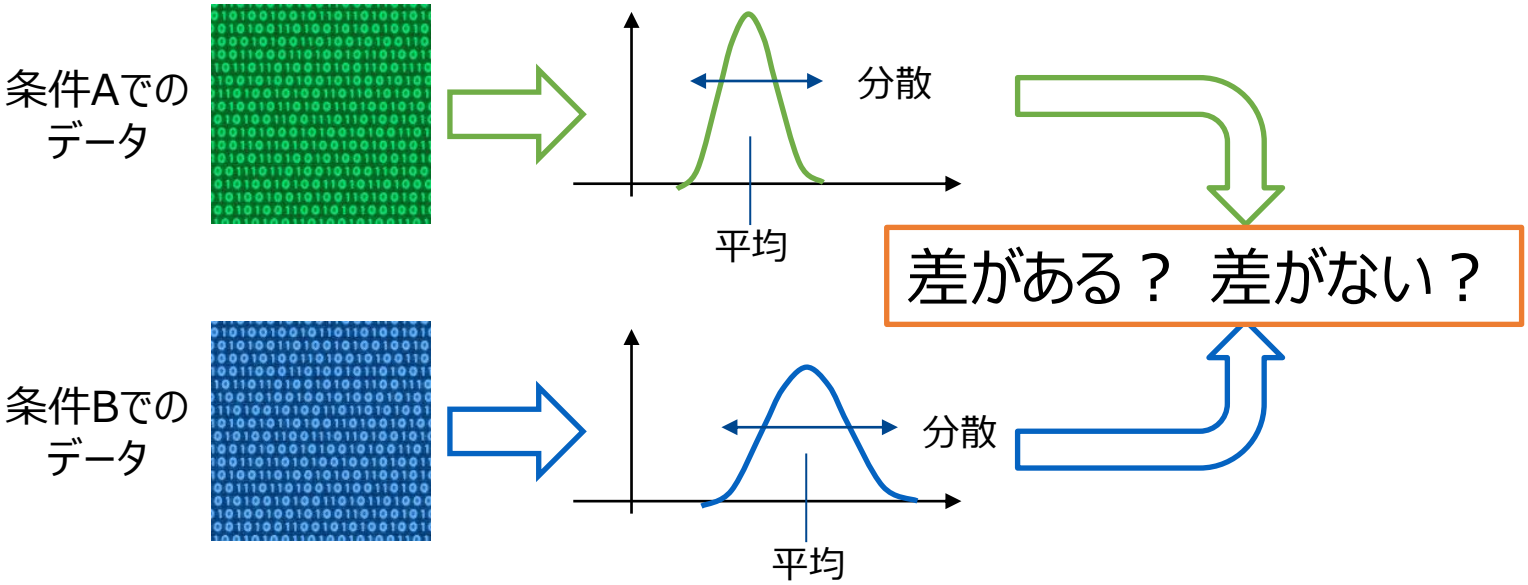


- ノンパラメトリックな確率分布
  - 任意形状の分布
  - 「○○分布」というような名前はない
  - ヒストグラムもその一種



確率分布がわかると何がウレシイのか？（1/2）

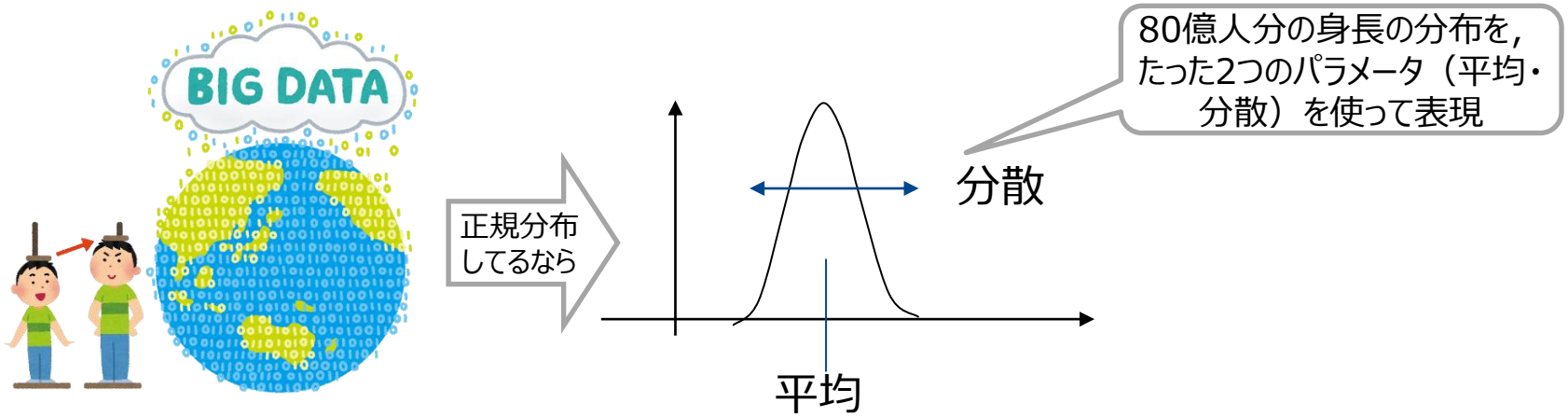
- 信頼区間や検定，回帰，認識など，様々な統計解析の基礎は，確率分布！



- 頻度に比べ，「合計＝1」に正規化されているので扱いやすい

確率分布がわかると何がウレシイのか？（2/2）

- さらにパラメトリックな分布なら…
  - データの分布を少数のパラメータを持つ「数式」で表せる
    - 多くのデータは，パラメトリックな確率分布に従うことが多い



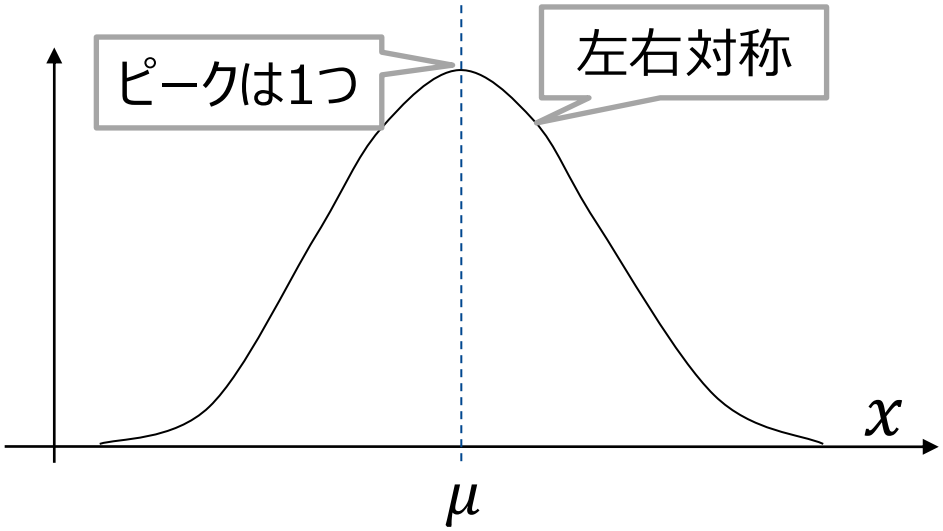
- では，数式で表されると，何がウレシイ？
  - 全データを覚えておかなくても，その数式(のパラメータ)があればOK
  - 様々な数学的演算(微分や積分など)を式の上で行えたり，比較が容易に

正規分布

「連続」なパラメトリック確率分布の代表格！
様々な講義で何度も出会うことになると思います

(1次元)正規分布:
ガウス分布とも呼ばれます

J.C.F. Gauss [1777-1855]



- 「連続」なパラメトリック確率分布の代表格！
  - パラメータは平均と分散の2つ
- 様々な講義で何度も出会うことになると思います

(1次元)正規分布:
式で書くと，思った以上に複雑に..!?

$\bar{x}$

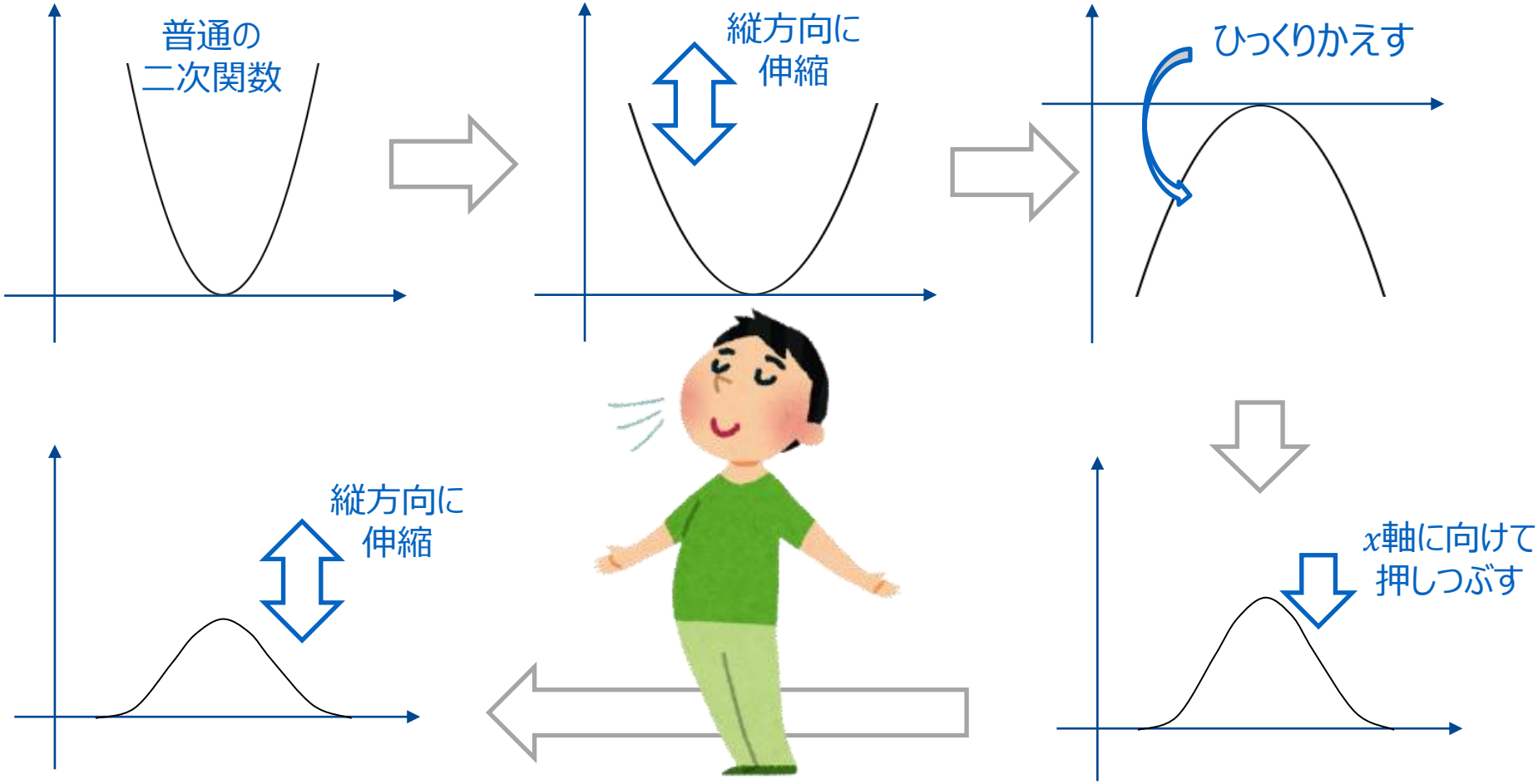
$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- σ^2: 分散
- μ: 平均(これまでは$\bar{x}$と書いてました)

- 複雑に見えますが，色々なところで出てくるので，
  これを機に落ち着いて考えてみよう！

そう，落ち着きましょう

- 順を追ってみていけば，そう難しい話でもないんです



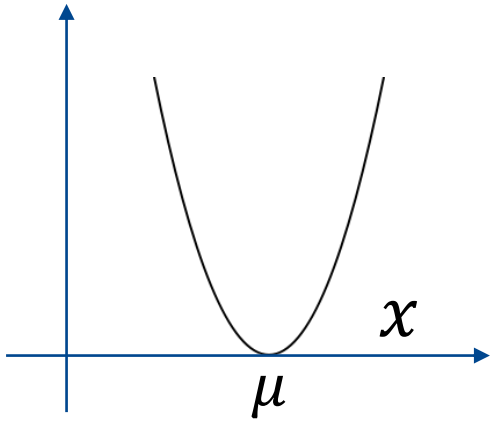
難しそうに見える正規分布，本質は単なる「二次関数」！(1/5)

- 大事なのは，「xが変わるとどう変化するか」だから，xの周りに注目

xに関する二次関数！

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- その部分だけグラフにすると



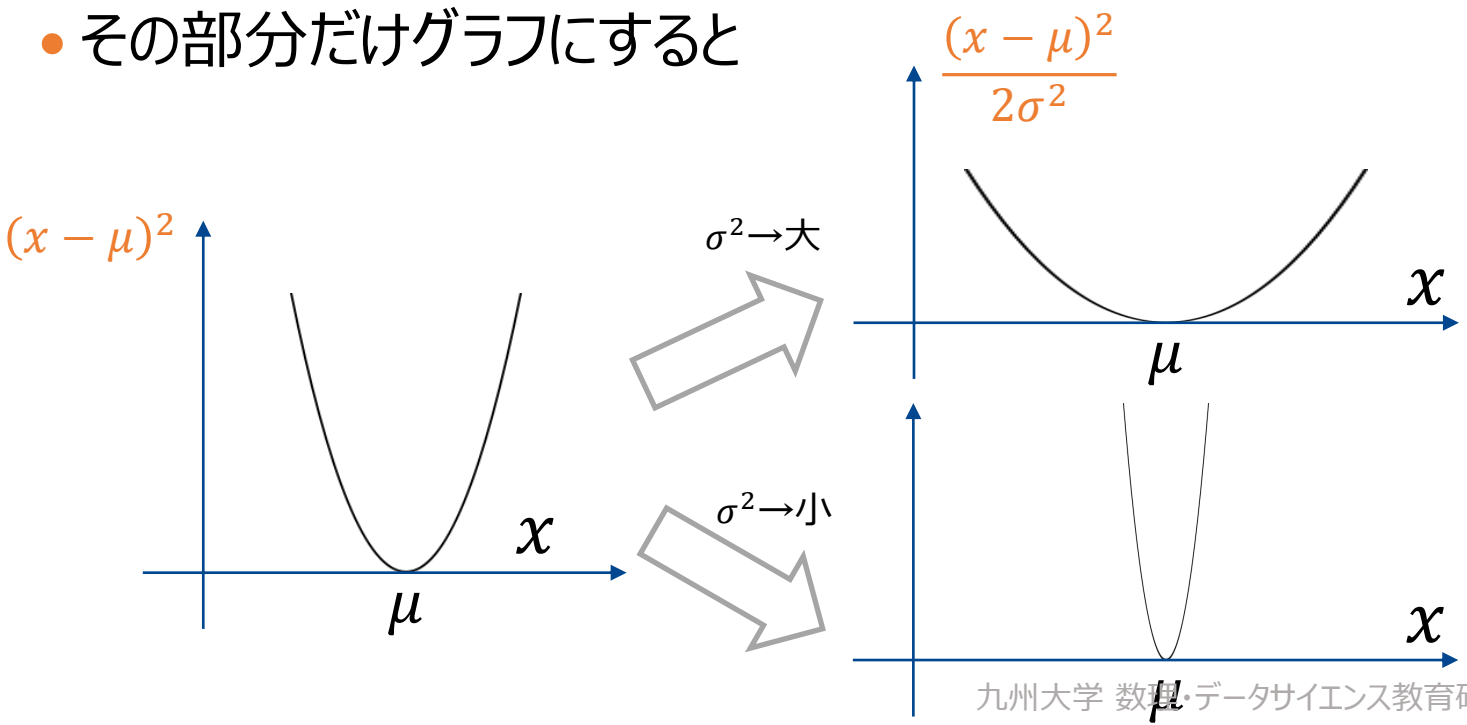
難しそうに見える正規分布，本質は単なる「二次関数」！(2/5)

- もうちょっと周りも見てみると

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

xに関係ない定数

- その部分だけグラフにすると



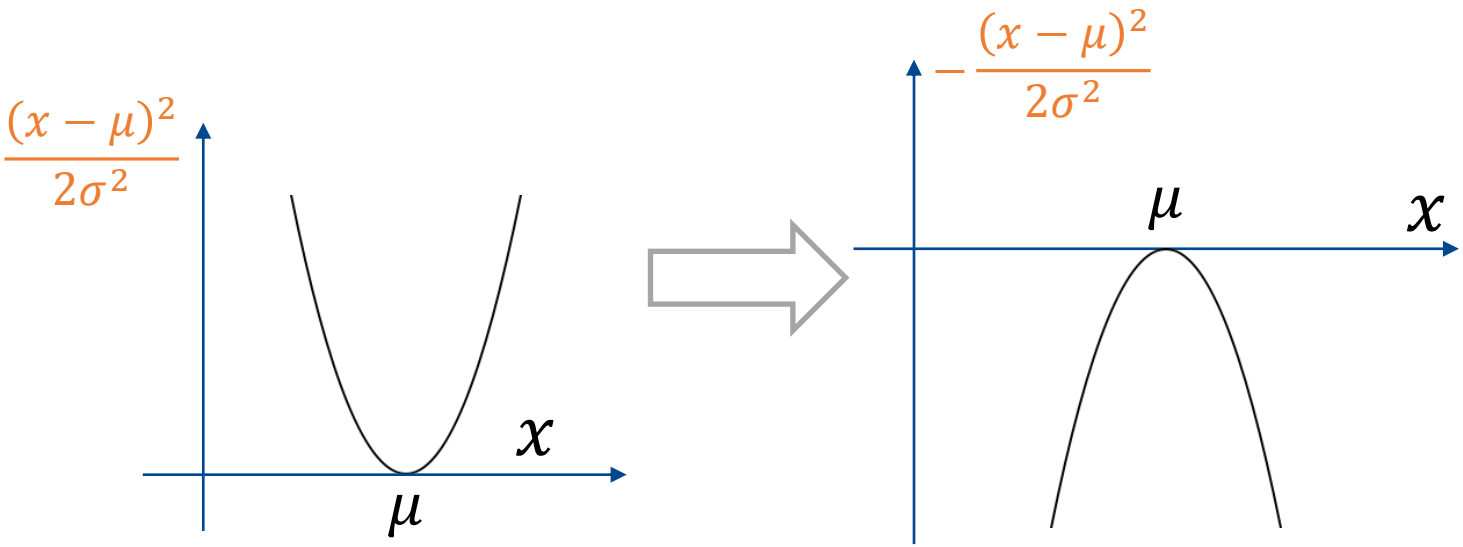
いずれにせよちょっと変形するだけで，相変わらず二次関数

難しそうに見える正規分布，本質は単なる「二次関数」！(3/5)

- もうちょっと周りも見てみると

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

- その部分だけグラフにすると

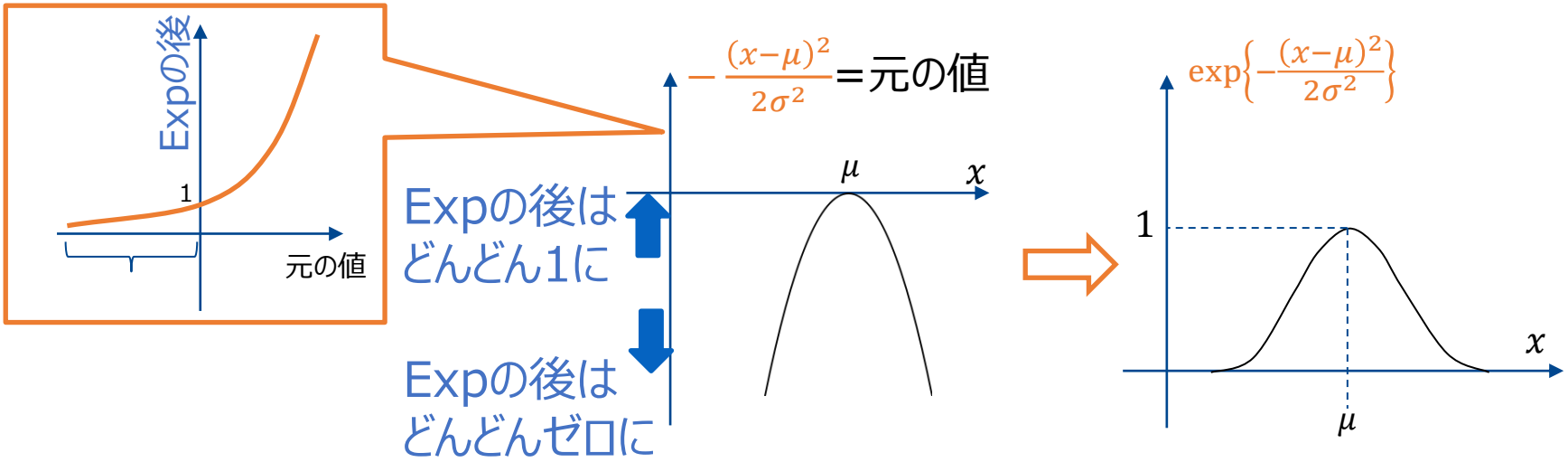


難しそうに見える正規分布，本質は単なる「二次関数」！(4/5)

- もうちょっと周りも見てみると

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- その部分だけグラフにすると

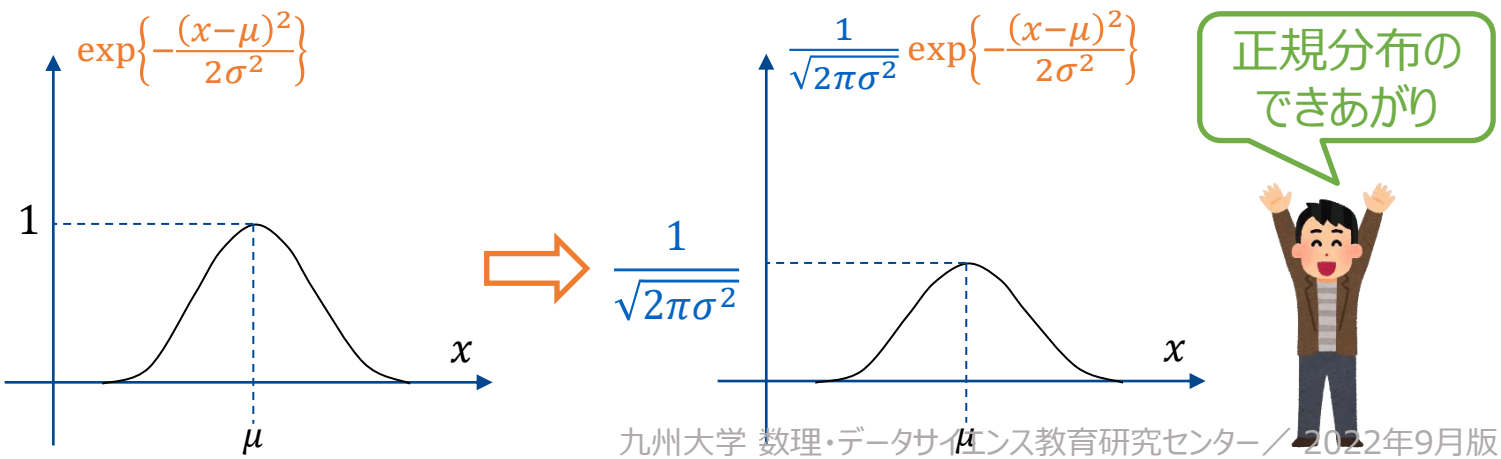


難しそうに見える正規分布，本質は単なる「二次関数」！(5/5)

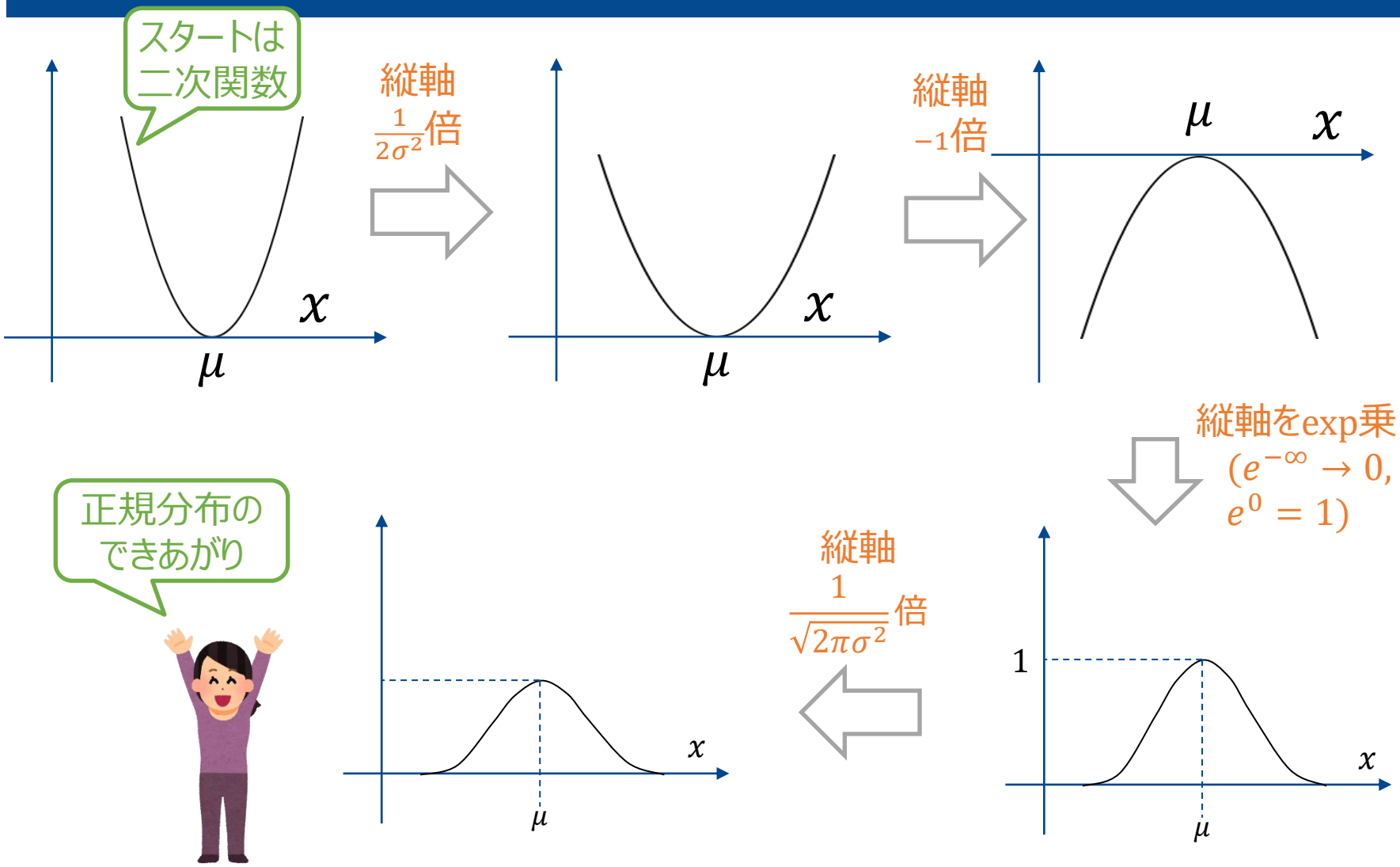
- もうちょっと周りも見てみると

$$p(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{x\text{に関係ない定数}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- 結局どうなるかというと



もう正規分布なんて怖くない！
二次関数の「縦軸」を何度かいじっただけ！

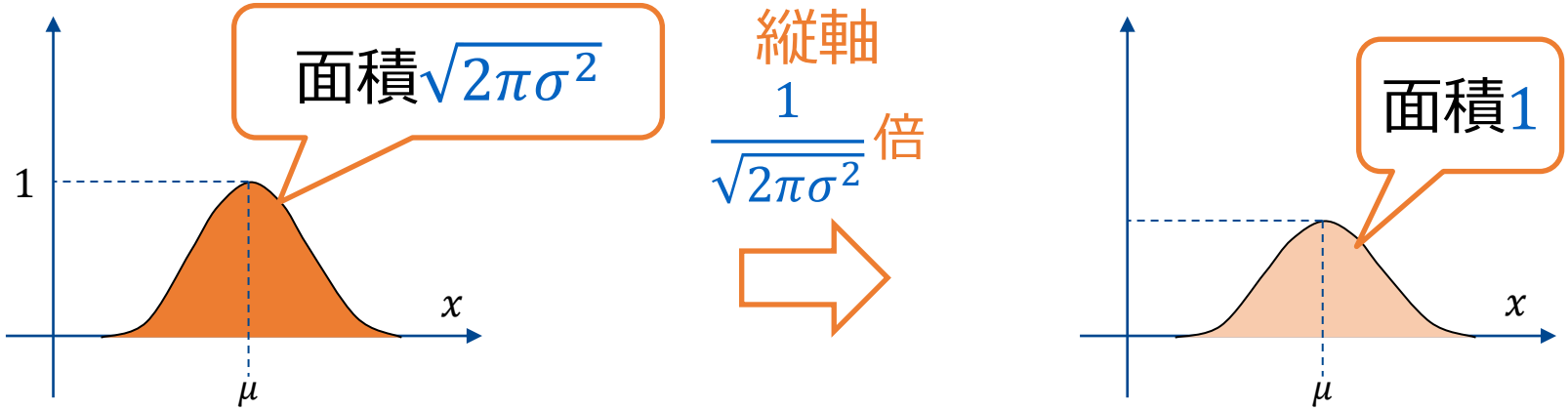


ちなみに，冒頭のゴチャゴチャ～っとしたところは「面積を1にするため」なので，気にしない

$$p(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

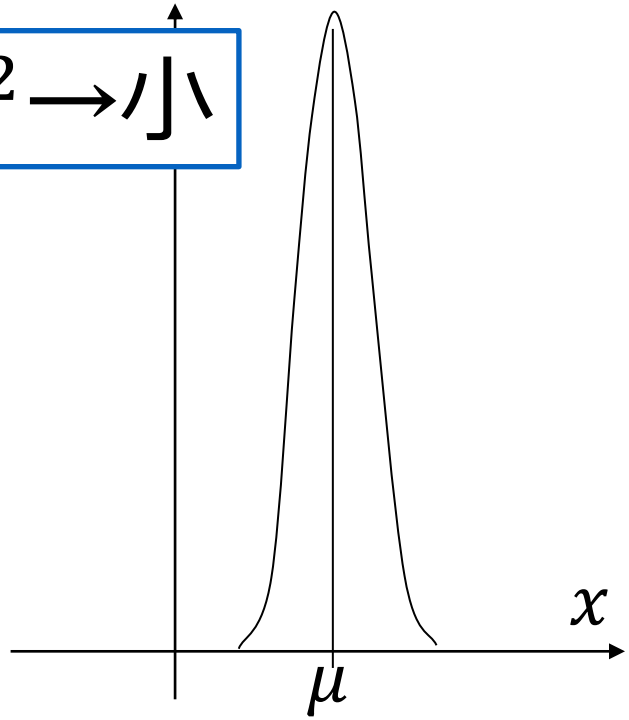
Q. このごちゃごちゃ～っとしたところはなぜ必要？

A. 確率密度関数とならば，面積が1となる必要があるため

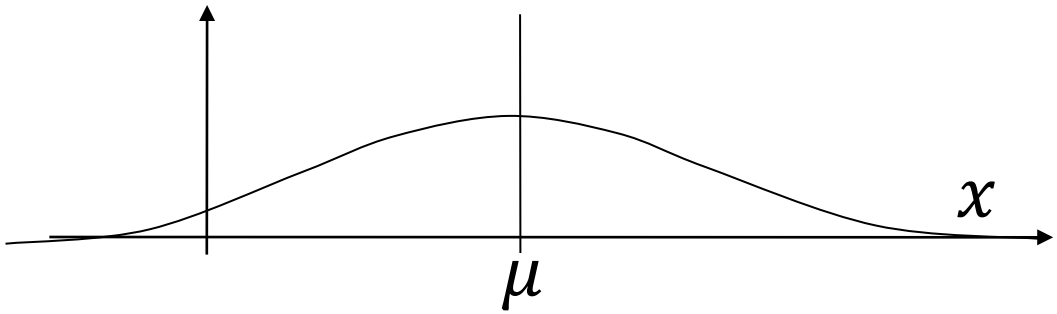


1次元正規分布の分散(1/2)

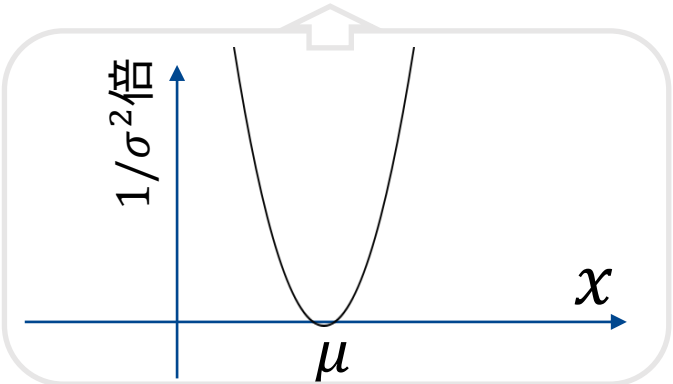
$\sigma^2 \to$ 小

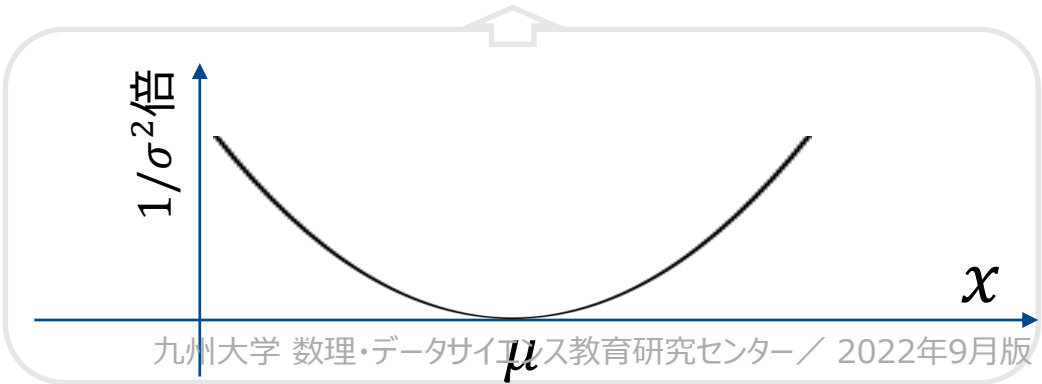


$\sigma^2 \to$ 大

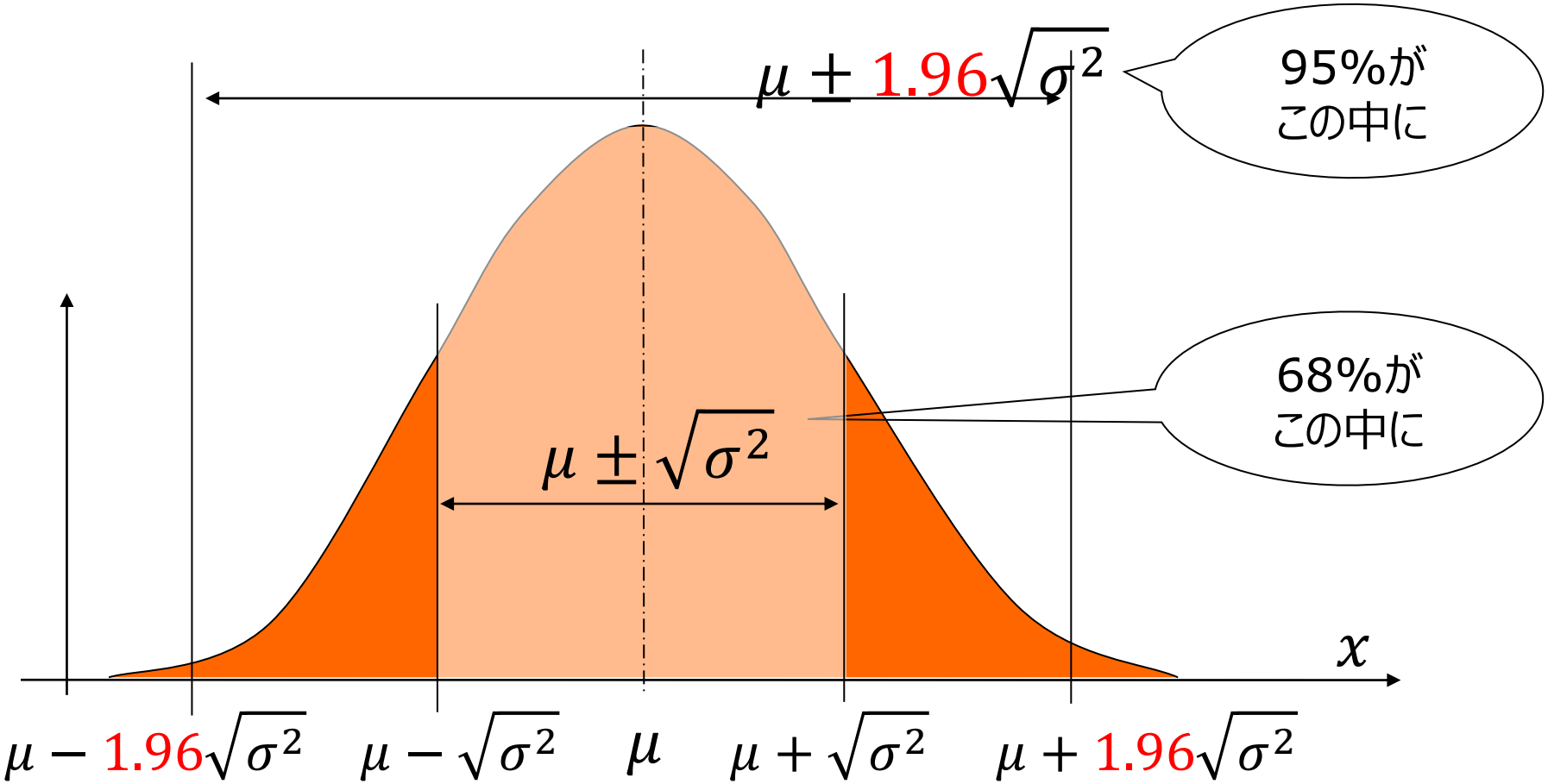


前身の二次関数





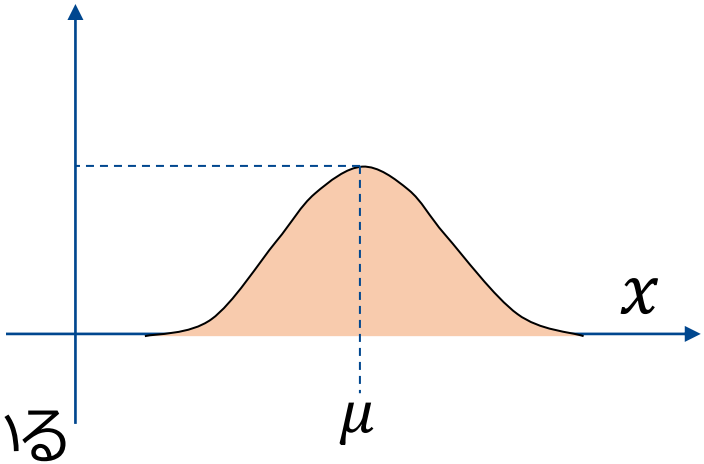
正規分布の分散(2/2)
「自分がどれぐらい外れているか」がわかる



$\mu \pm 2.57\sqrt{\sigma^2} \rightarrow 99\%$

式はわからなくてもＯＫ．それより大事な「正規分布の性質」のまとめ

- 平均付近の値が一番出やすい
- 平均から離れた値ほど出にくくなる
  - 出にくくなるスピードは分散に関係
- 平均を中心に左右対称
- すそ野は，$x = -\infty$から$+\infty$まで広がっている
  - 出やすさはゼロに限りなく近いがゼロではない

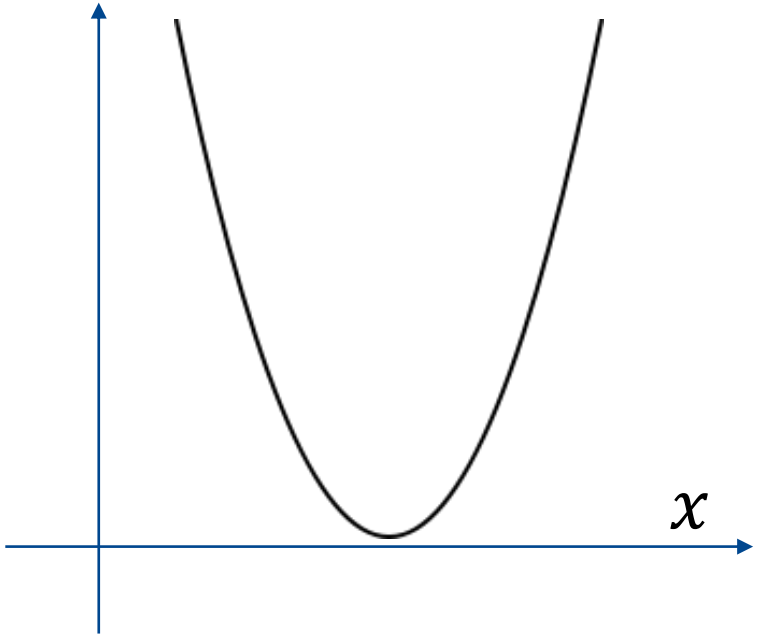


多次元正規分布

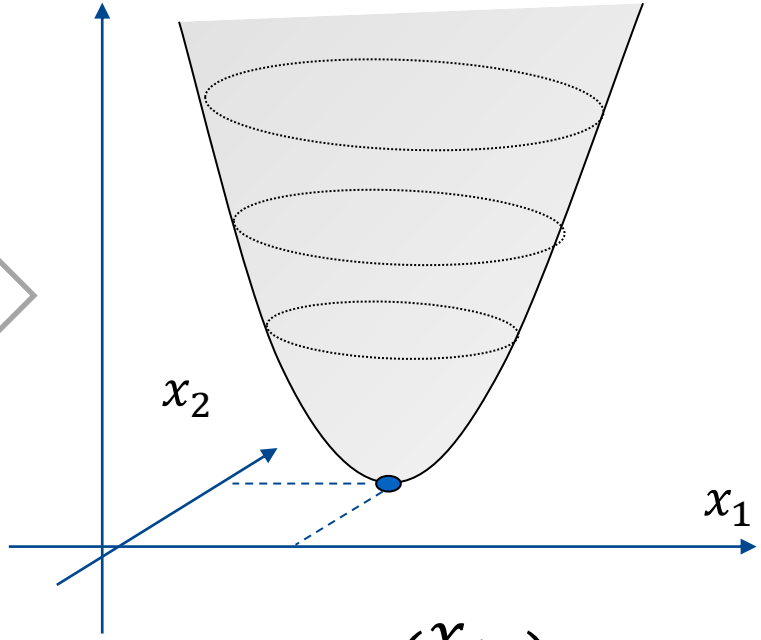
ベクトルだって正規分布する

ちょっと準備．d次元ベクトルに対する2次関数 (1/2)

- $d = 2$の場合

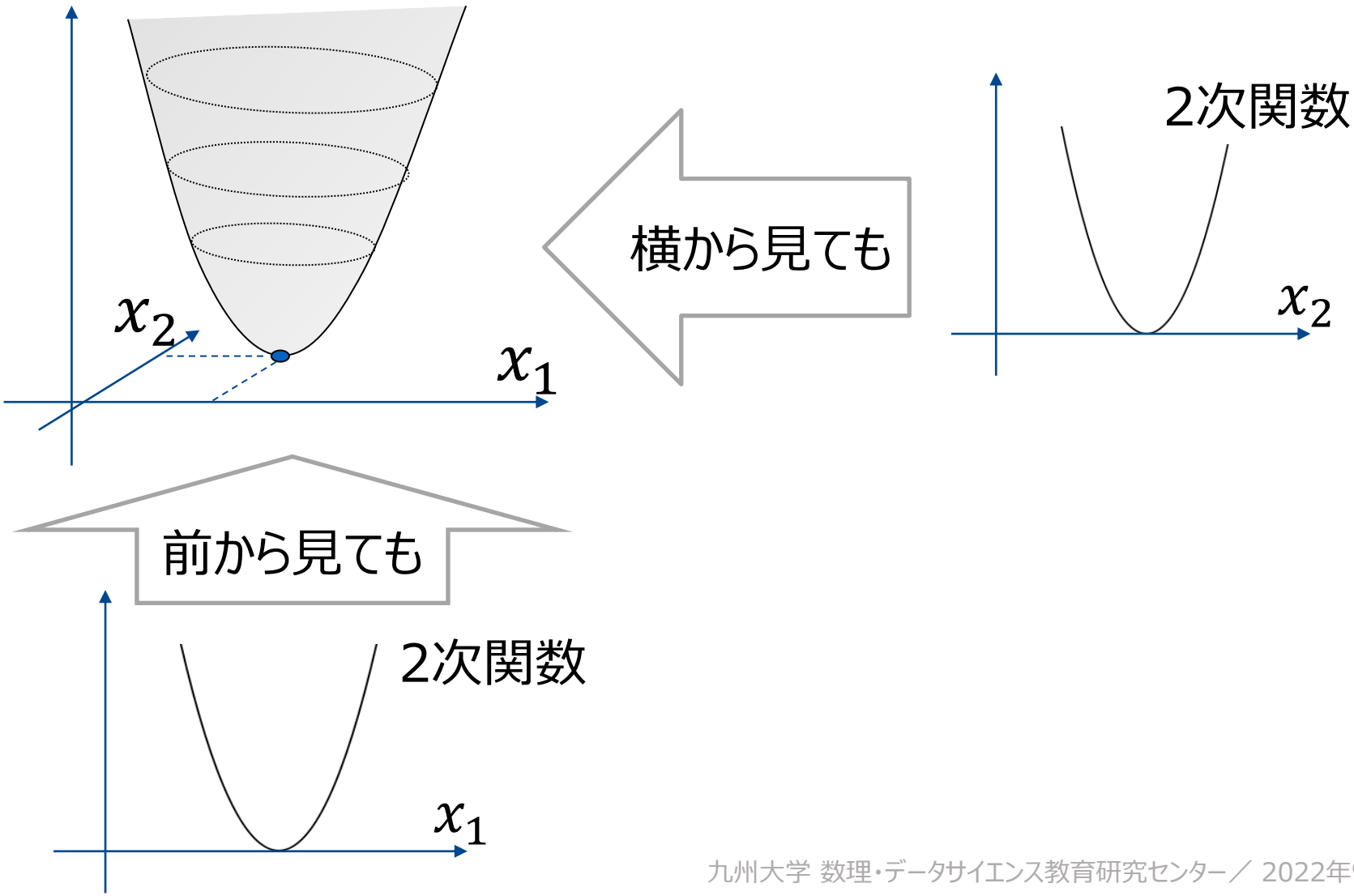


$x =$ 数値

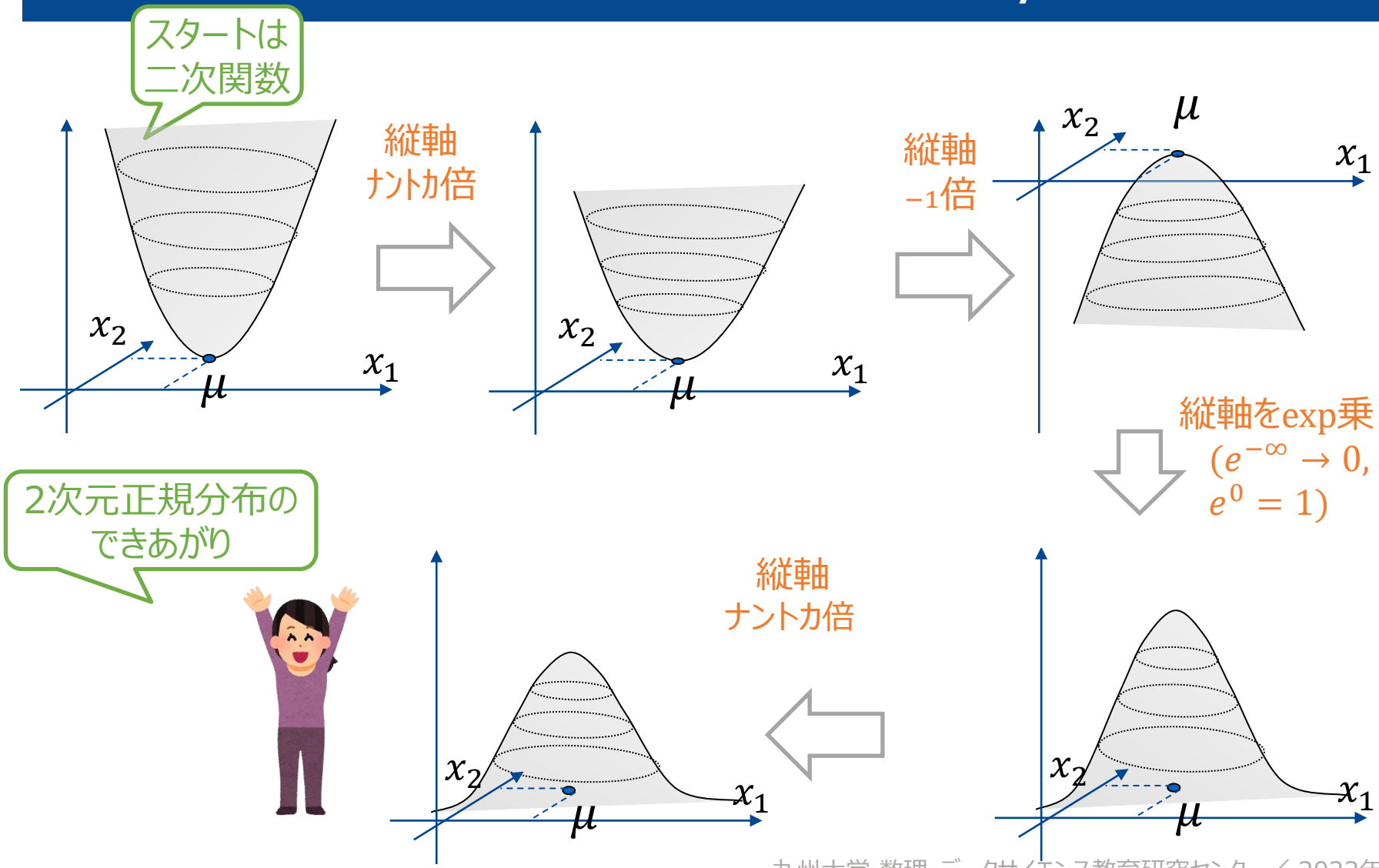


$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

ちょっと準備．d次元ベクトルに対する2次関数 (2/2)

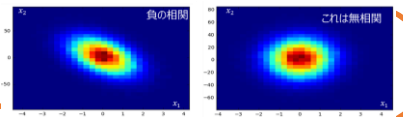


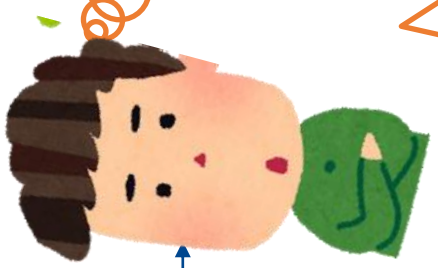
ということは，2次元の場合も1次元と同じような理屈が成り立つのでは？→yes!

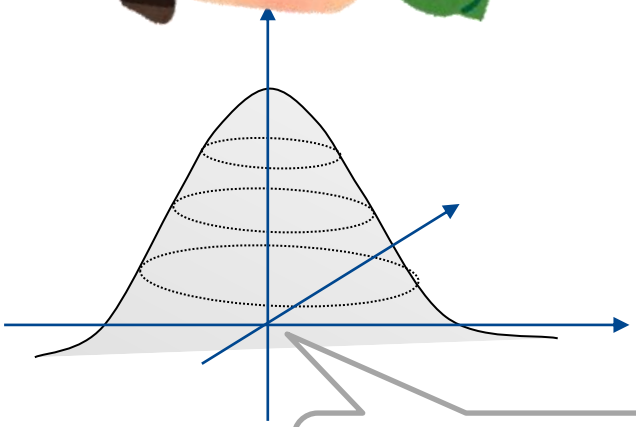


1次元と２次元の違い：分布の広がり方がちょっと複雑に

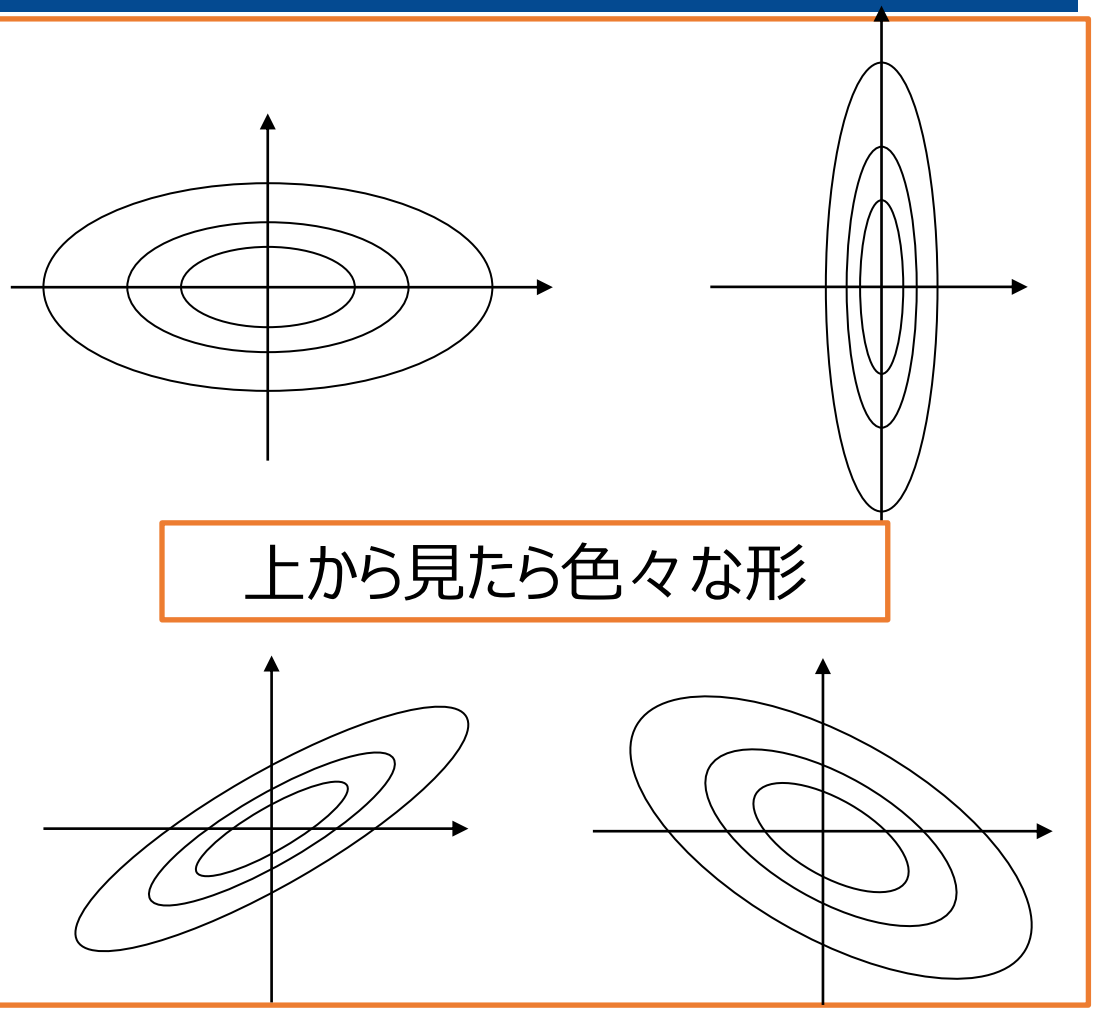
あ，そうか．相関か







面倒なので
平均＝ゼロで説明



d次元正規分布，式で書くとクラクラする!

d次元ベクトル

$d \times d$「行列」

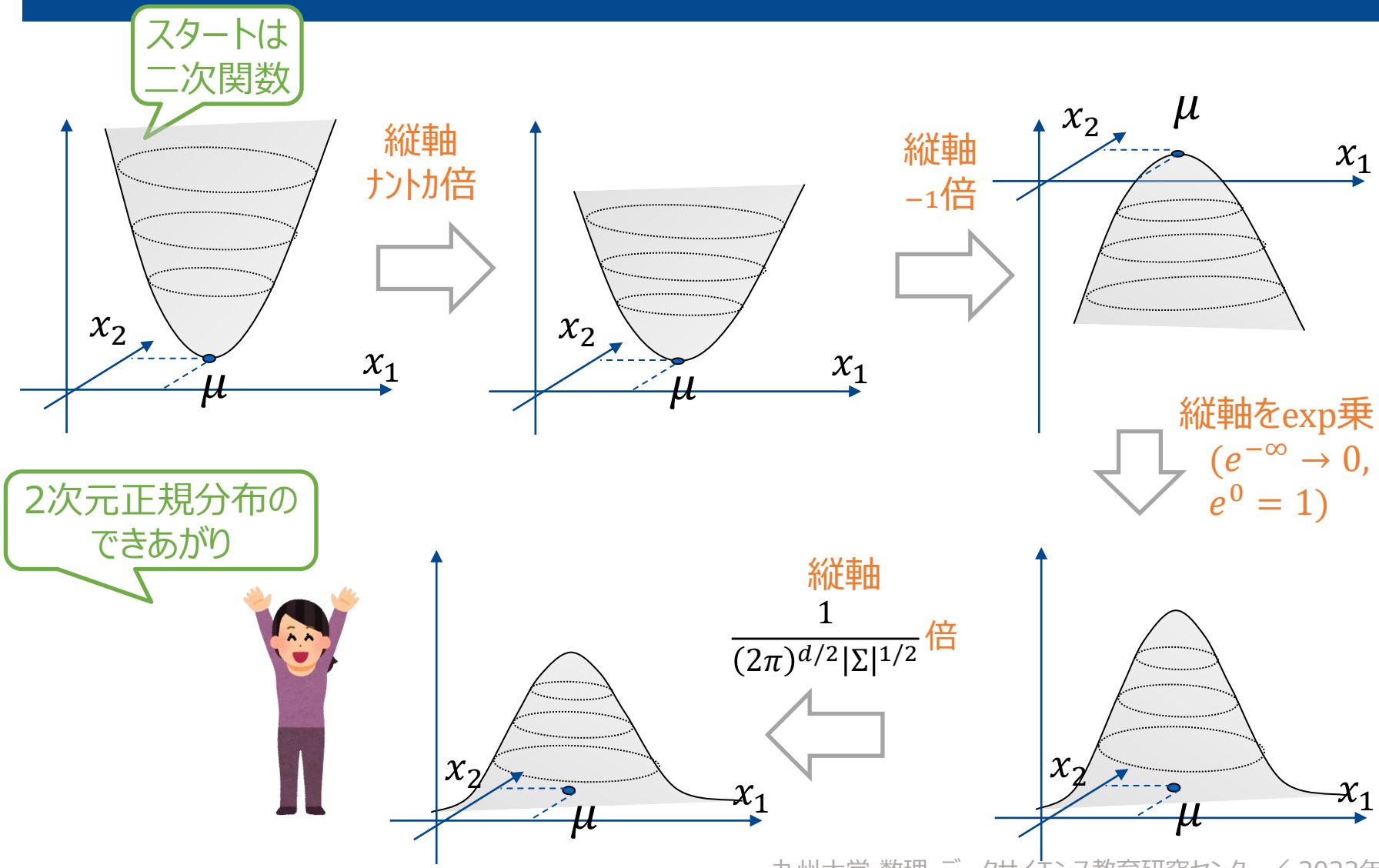
$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

- $\boldsymbol{\Sigma}$: 共分散行列
- $\boldsymbol{\mu}$: (d次元)平均ベクトル

大文字のシグマΣを使う.
総和記号と混同しないように

- 1次元で本質は簡単だったことを見たので，
今回もやはり落ち着いて考えてみよう！

d次元の場合も1次元と同様！
以下は$d = 2$の場合



そうなのかもしれないが，やっぱり「行列」のところが気になる

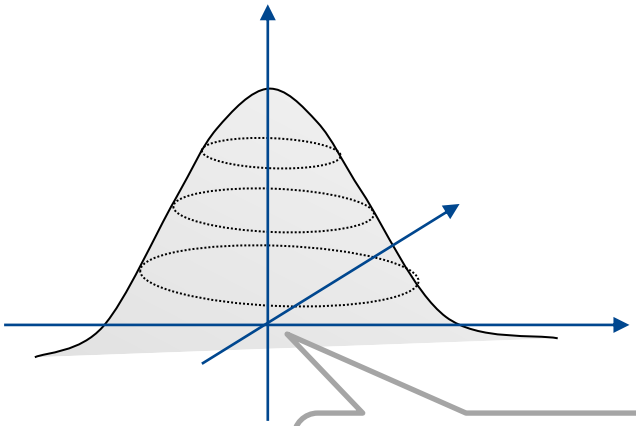
$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

- $\boldsymbol{\Sigma}$：共分散行列

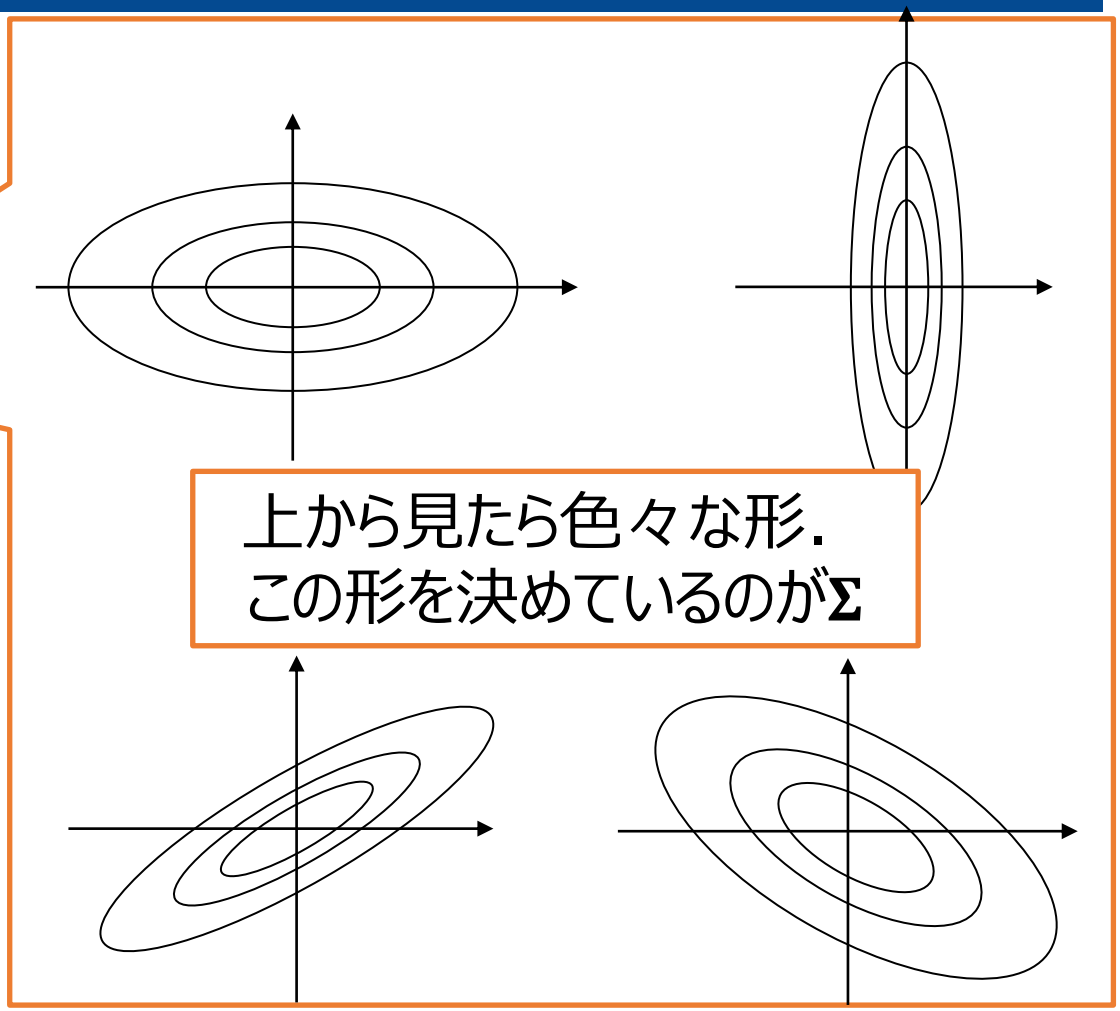
何者なんだ？
嫌がらせか？

共分散行列Σの役目を言ってしまうと…

- 分布の広がり方を決めている！



面倒なので
平均＝ゼロで説明

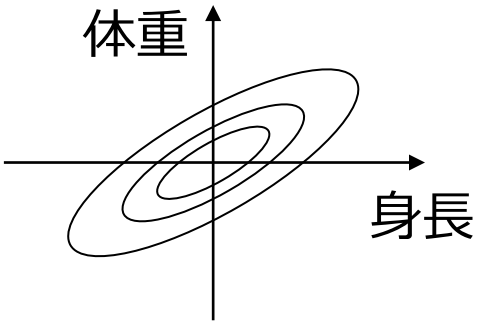


上から見たら色々な形.
この形を決めているのがΣ

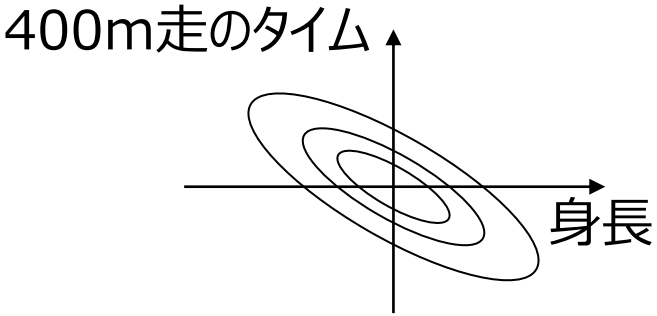
これがわかっておけばOK

分布の広がり方(傾き方)と相関

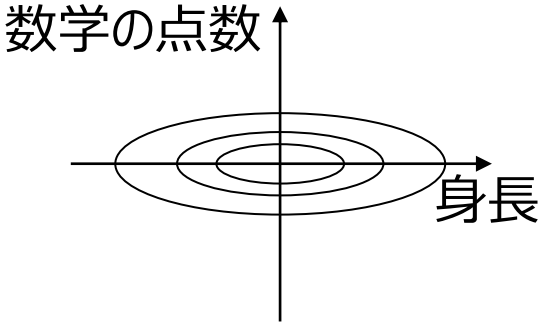
- x_1 が大ならx_2も大(＝正の相関)
  - 身長と体重



- x_1が大ならx_2は小(＝負の相関)
  - 身長と400m走のタイム



- x_1とx_2は無関係(＝無相関)
  - 身長と数学の点数



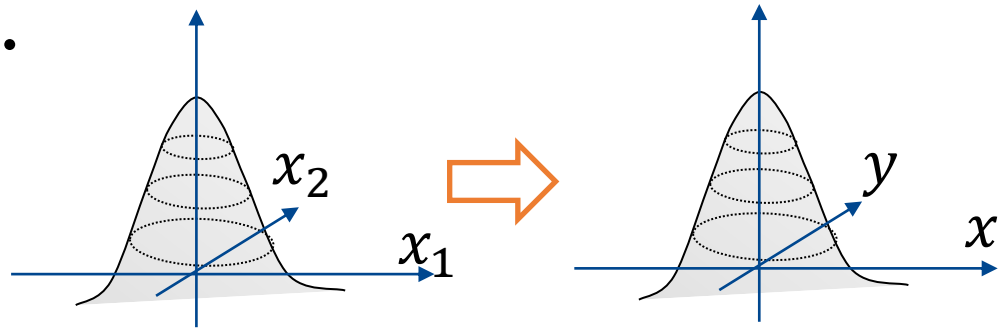
共分散行列の詳細は付録

【付録1】
共分散行列の詳細

避けては通れない行列.

d =2次元の場合の共分散行列を見てみよう (1/8)

- 添え字がごちゃごちゃするのが嫌なので$\boldsymbol{x} = (x_1, x_2)$を$\boldsymbol{x} = (x, y)$と書くと…



$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

d =2次元の場合の共分散行列を見てみよう (2/8)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

行列式 $|\boldsymbol{\Sigma}|$

$$= \sigma_x^2 \sigma_y^2 - \left(\sigma_{xy}\right)^2$$

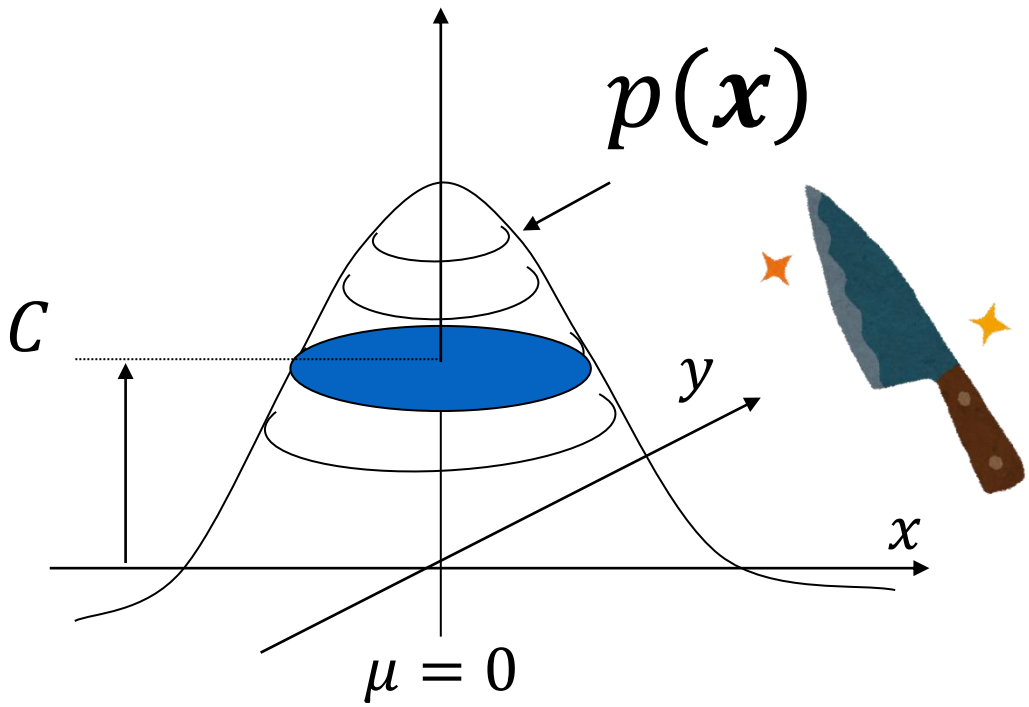
逆行列 $\boldsymbol{\Sigma}^{-1}$

$$= \frac{1}{|\boldsymbol{\Sigma}|} \begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix}$$

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

d =2次元の場合の共分散行列を見てみよう (3/8)

- 分布の形を見るため，その断面を見てみる



d =2次元の場合の共分散行列を見てみよう (4/8)

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}=C$$

単なる定数，両辺をこれで割ると

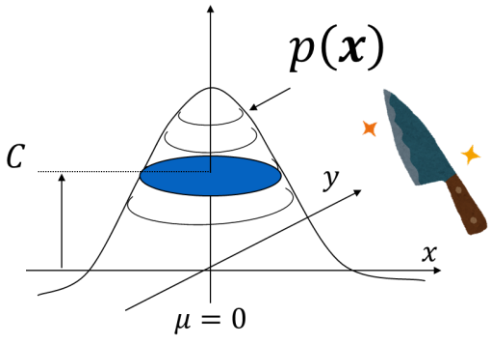
$$\Leftrightarrow \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}=C'$$

両辺をlogとって-2倍すると

$$\Leftrightarrow (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=C''$$

いま平均$\boldsymbol{\mu}$はゼロとしているので

$$\Leftrightarrow \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}=C'''$$



d =2次元の場合の共分散行列を見てみよう (5/8)

$$\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x} = C''' \Leftrightarrow \boldsymbol{x}^T \frac{1}{|\boldsymbol{\Sigma}|} \begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix} \boldsymbol{x} = C'''$$

$$\Leftrightarrow (\boldsymbol{x} \quad \boldsymbol{y}) \begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = C''''$$

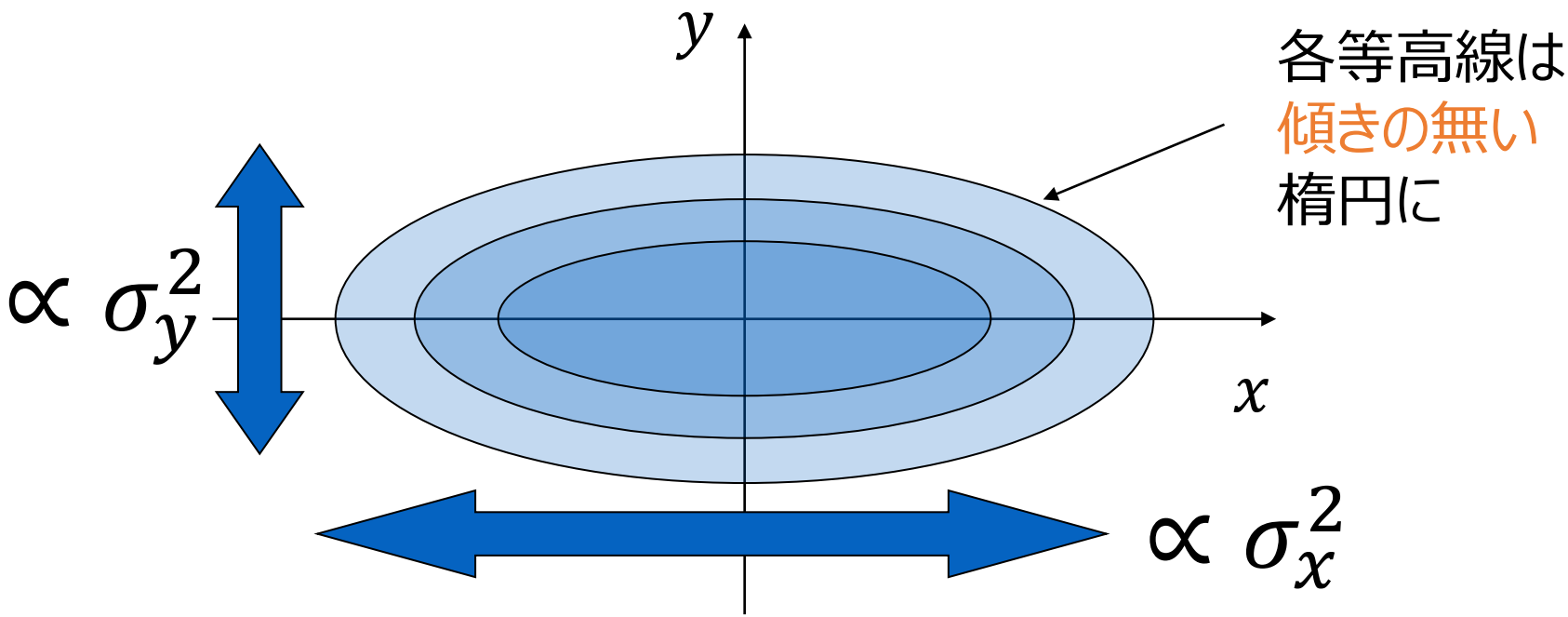
楕円の方程式！

$$\Leftrightarrow \boxed{\frac{1}{\sigma_x^2} x^2 - 2\frac{\sigma_{xy}}{\sigma_x^2 \sigma_y^2} xy + \frac{1}{\sigma_y^2} y^2 = \text{定数}}$$

d =2次元の場合の共分散行列を見てみよう (6/8)

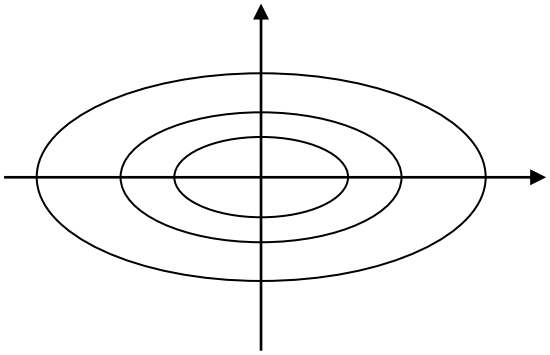
- $\sigma_{xy} = 0$だったら…

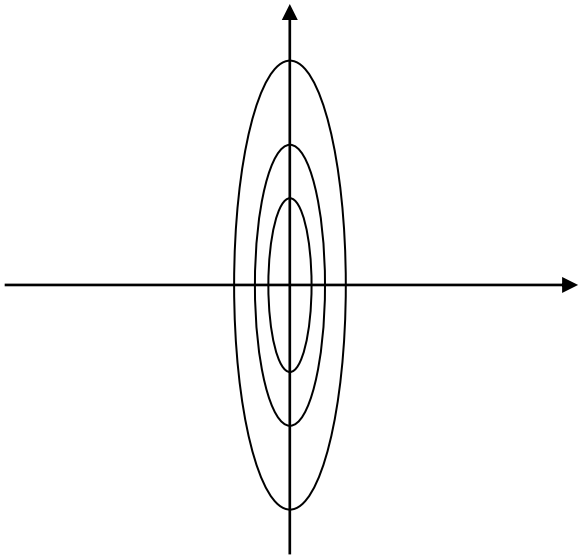
$$\frac{1}{\sigma_x^2}x^2 + \frac{1}{\sigma_y^2}y^2 = 定数$$



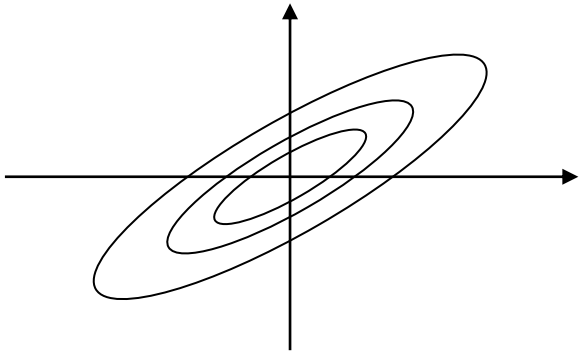
d =2次元の場合の共分散行列を見てみよう (7/8)

- $\sigma_{xy} = 0$： 傾きなし





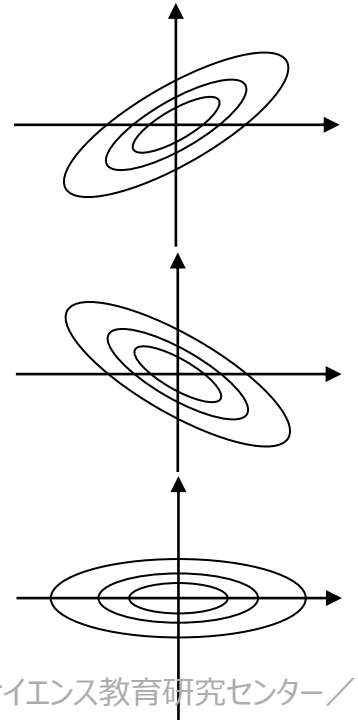
- $\sigma_{xy} \neq 0$： 傾きあり



d =2次元の場合の共分散行列を見てみよう (8/8)

- 対角成分：σ_x^2, σ_y^2
  - 各軸方向の広がり（分散）

- 非対角成分：σ_{xy}
  - 2軸間の相関を表す
  - $\sigma_{xy} > 0$: x が大なら y も大 → 正の相関
  - $\sigma_{xy} < 0$: x が大なら y は小 → 負の相関
  - $\sigma_{xy} = 0$: xと y は無関係(独立) → 無相関

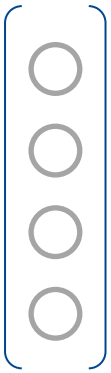


【付録2】
というか，行列って何？

並ぶとできる「行列」ではない

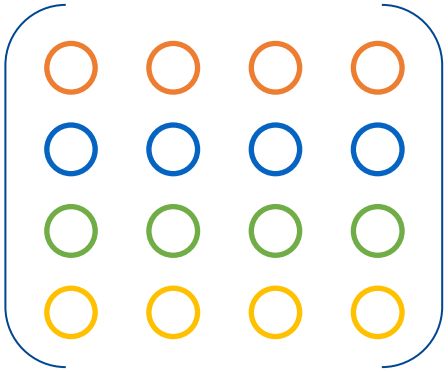
行列って…
数字が縦横に並んだものです

ベクトル



数字が一列に並ぶ

行列

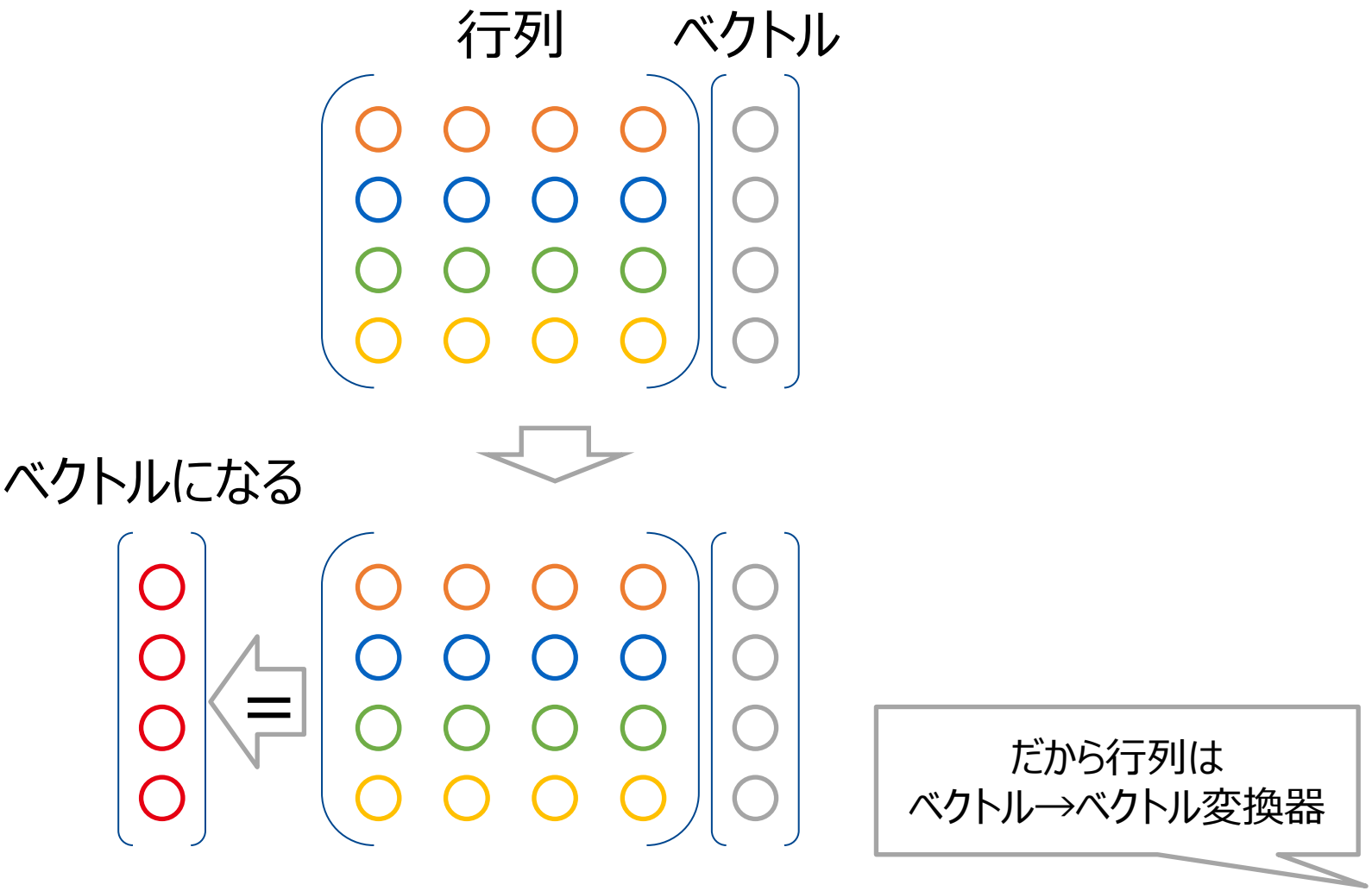


数字が縦横に並ぶ

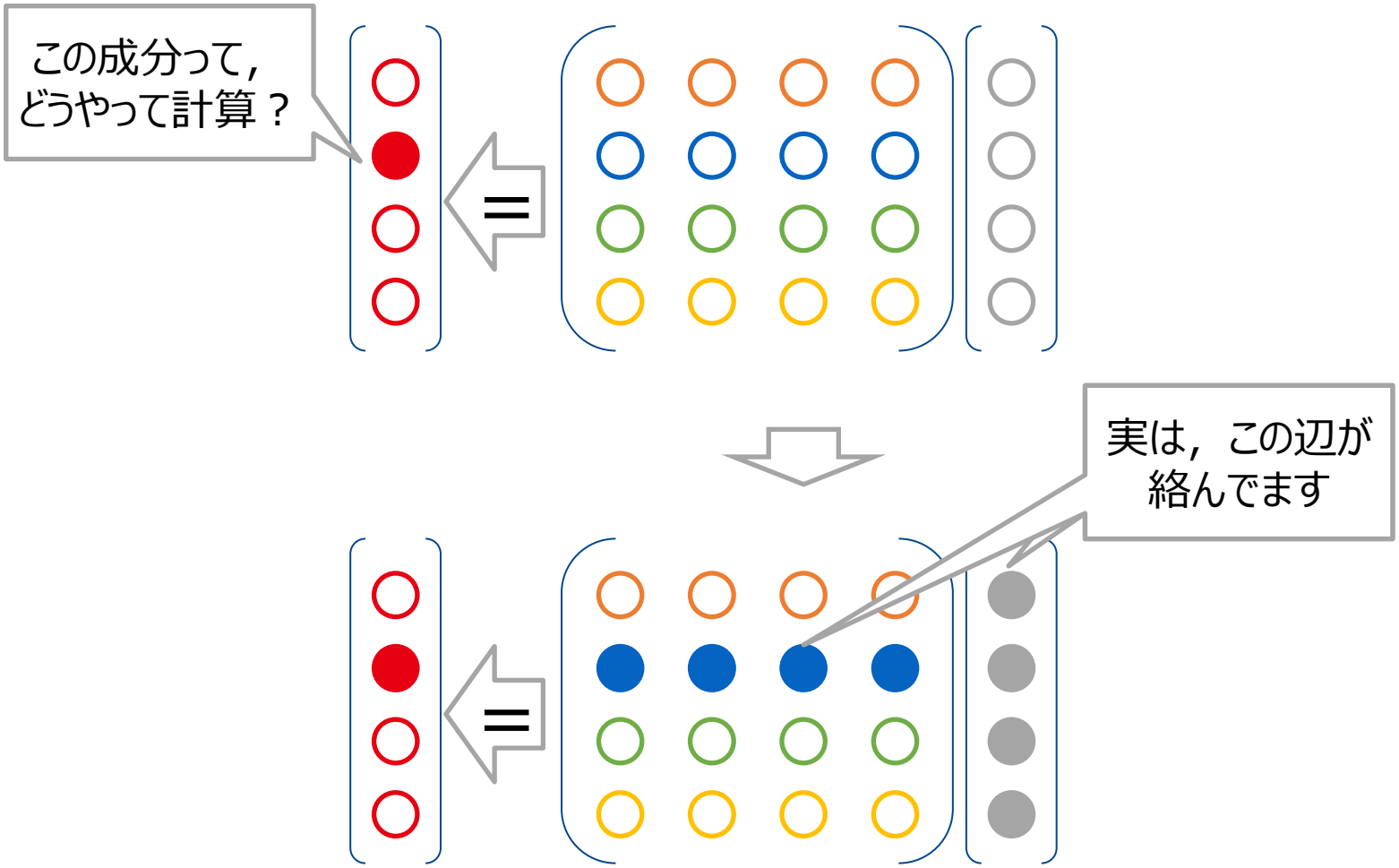
2×2行列

例えば：$\begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}$

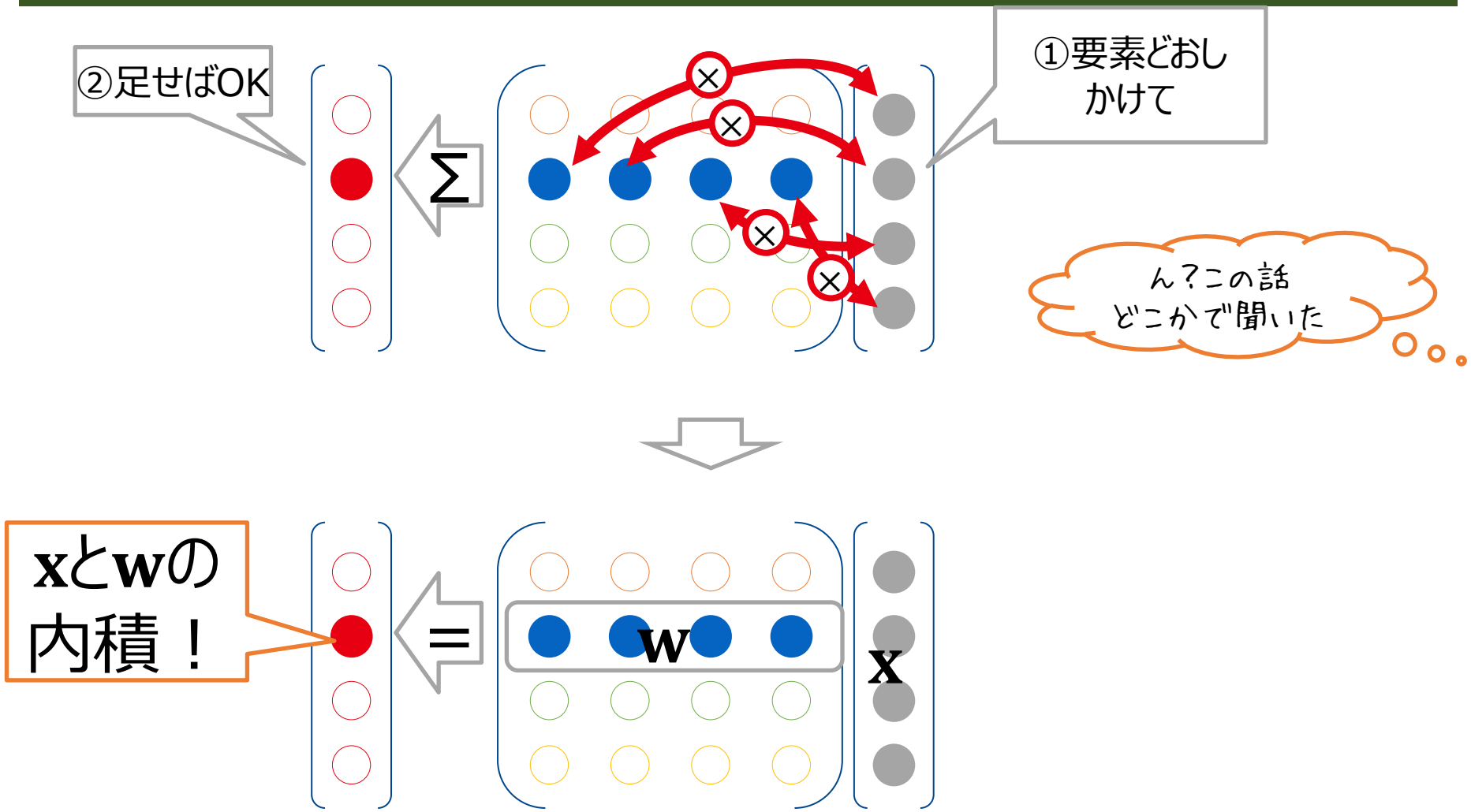
行列とベクトルの意外(?)関係：行列とベクトルの積…(1/6)



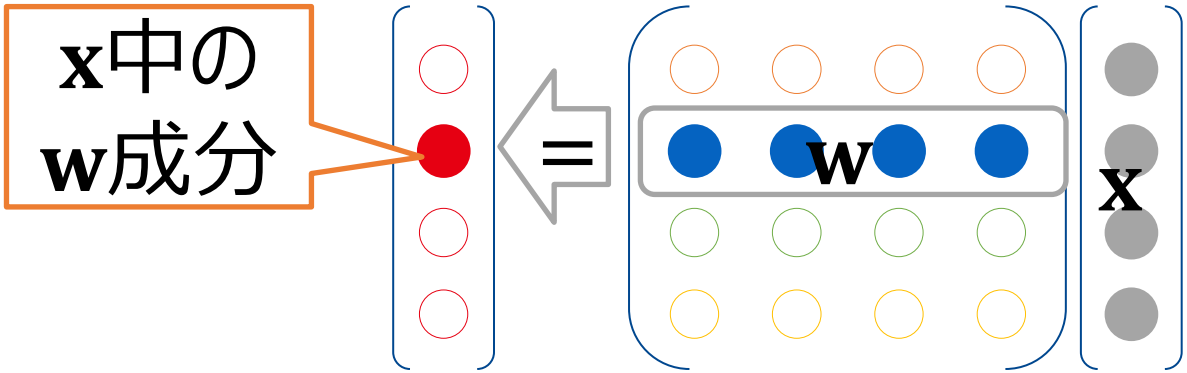
行列とベクトルの意外(?)関係：行列とベクトルの積…(2/6)

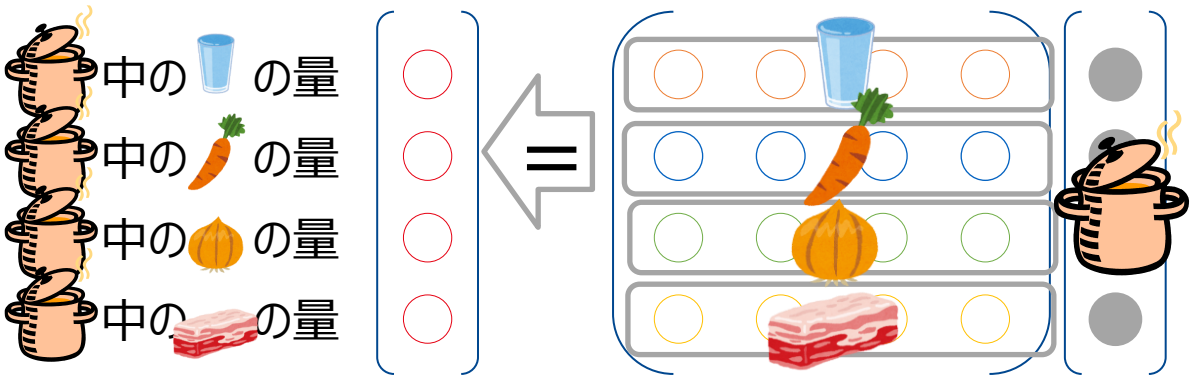


行列とベクトルの意外(?)関係：行列とベクトルの積…(3/6)



行列とベクトルの意外(?)関係：行列とベクトルの積…(4/6)

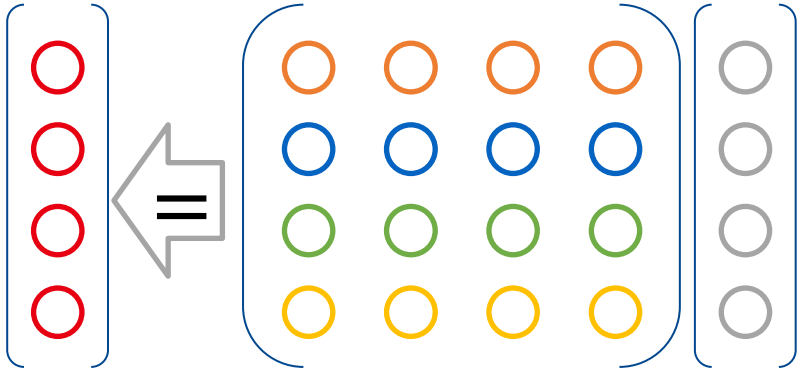




行列は一括分析器でもある！
様々な成分を一度に抽出

「だったら，行列って基底ベクトルが
並んでるものと考えられる？」と
思ったアナタは素晴らしい

行列とベクトルの意外(?)関係：行列とベクトルの積…(5/6)



行列は
ベクトル→ベクトル変換器

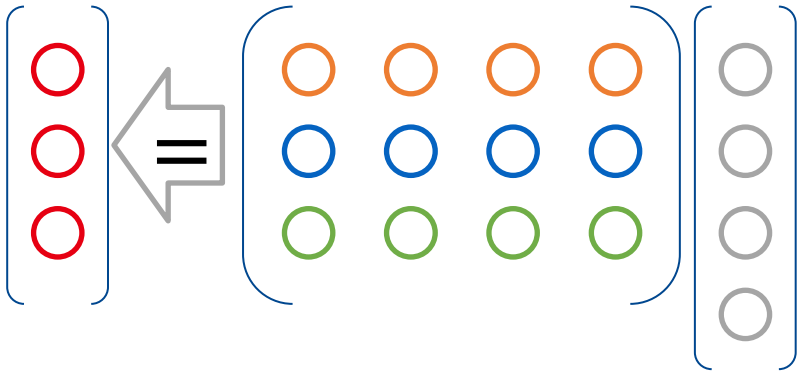
行列は
一括分析器！

Don’t worry. 同じもの.
変換結果は分析結果としても見れる，というだけ

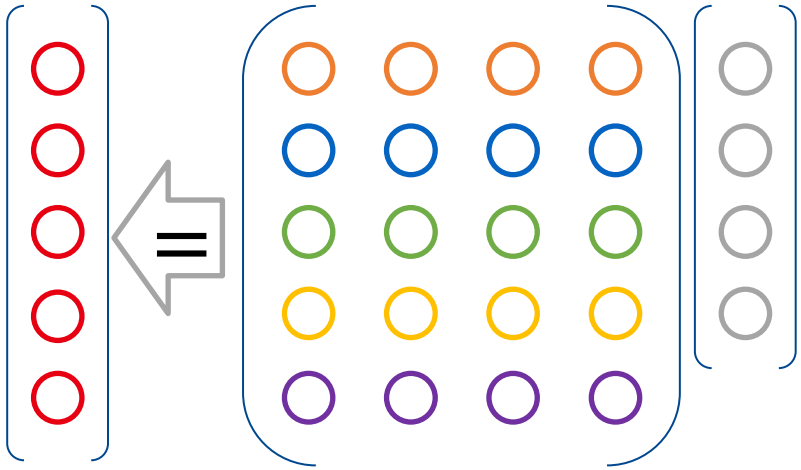
行列とベクトルの意外(?)関係：行列とベクトルの積…(6/6)

- 変換により次元を変えちゃうこともできる!

次元が下がった



次元が上がった



というわけで「行列」

- だれかが並んでいるわけではない

- ベクトルを別のベクトルに変換するのに使える

- 他にも色々な用途がある！
  - 画像を変形するのも「行列」
  - AI（深層ニューラルネットワーク）の中身も「行列」
  - 要素間の関係や計量(?)を与えるのも「行列」
  - 連立方程式も「行列」 $\begin{cases} 2x + 3y = 6 \\ -x + 5y = 10 \end{cases} \rightarrow \begin{bmatrix} 2 & 3 \\ -1 & 5 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 6 \\ 10 \end{pmatrix}$
  - 興味を持ったら線形代数を勉強してみましょう！

【付録3】
あの2はどこから出てくるのか？

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

あの「２」はどこから出てくるのか？

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- 正規分布が満たすべき条件から導出されます

簡単のため$\mu = 0$とし，$p(x) = a\exp(-x^2/b)$として，条件

$$\begin{cases} \int p(x)dx = 1 \\ \int x^2 p(x)dx = \sigma^2 \end{cases}$$

を満たすa, bを求めると，$\mathrm{a} = \frac{1}{\sqrt{2\pi\sigma^2}}, b = 2\sigma^2$となる．