機械学習・ディープラーニングのための

基礎数学講座　確率・統計 Day 3

SkillUP AI

配布物

- 1_slide：スライド教材が入ったフォルダ
  - prob_and_stats_DAY3.pdf：このスライド
- revision_history.txt：改訂履歴

参考文献（DAY1～DAY3を通して）

- **増補改訂版 語りかける中学数学**
  - https://www.beret.co.jp/books/detail/459
  - 中学数学が怪しい方へ
- **ライブ講義　大学１年生のための数学入門**
  - https://bookclub.kodansha.co.jp/product?item=0000275978
  - 数学で出てくる記号の意味が怪しい方へ
- **流れるようにわかる統計学**
  - https://www.kadokawa.co.jp/product/301312000211/
- **確率統計キャンパス・ゼミ 改訂6**
  - https://www.mathema.jp/product/確率統計キャンパス・ゼミ-改訂6/
- **演習 確率統計キャンパス・ゼミ 改訂4**
  - https://www.mathema.jp/product/演習確率統計キャンパス・ゼミ-改訂4/
  - 問題集形式

参考文献（DAY1～DAY3を通して）

- まなびのずかんシリーズ統計学の図鑑
  - https://gihyo.jp/book/2015/978-4-7741-7331-3
- Pythonで理解する統計解析の基礎
  - https://gihyo.jp/book/2018/978-4-297-10049-0
- なるほど統計学
  - https://www.amazon.co.jp/dp/4875252102
- 日本統計学会公式認定 統計検定 3級・4級 公式問題集
  - https://jitsumu.hondana.jp/book/b496705.html
  - 問題集形式
- データサイエンスのための統計学入門 第2版
  - https://www.oreilly.co.jp/books/9784873119267/
- 最短コースでわかる ディープラーニングの数学
  - https://www.nikkeibp.co.jp/atclpubmkt/book/19/273470/

本講座の全体の内容

Day 1 ：記述統計学の基礎

- 内容：統計量・可視化
- 修了演習：データの前処理技術（正規化と標準化）・箱ひげ図を用いた外れ値検出

Day 2 ：確率

- 内容：確率の基礎・条件付き確率・ベイズの定理・独立
- 修了演習：ナイーブベイズによるスパムメール判定

Day 3 ：確率分布

- 内容：離散型/連続型確率分布
- 修了演習：ロジスティック回帰

本講座でやること / やらないこと

- やること
  - 確率・統計分野の重要な概念・公式
  - 各種公式を使った問題演習
  - 機械学習 / 深層学習における上記概念・公式の利用方法の概要
- やらないこと
  - 紹介する公式等の厳密な証明
  - Python等を用いた実装方法
  - 機械学習 / 深層学習の各種手法の詳細な説明

講座に入る前に

- 青字・下線付きは URL リンク付き文字です
  - PDFビューワ上で該当箇所をクリックすると
    参考ページに遷移することができます
  - 例）スキルアップAI
    （スキルアップAIのトップページ https://www.skillupai.com/ へ遷移）

講座に入る前に

- 本講座では機械学習 / 深層学習を学ぶための土台となる内容を学習します
  そのため、目的意識を持って学び、アウトプットすることが重要です
- そこで、次の2点を必ず実施しましょう

1. **事前にスライドに目を通し、予習を行いましょう**
   - 漫然と目を通すだけでなく、どの部分を集中して聞くべきか自分の中で決めておきましょう
2. **各 DAY ごとに振り返り・言語化の時間を取りましょう**
   - 振り返り内容
     - この DAY で学んだ内容で参考になったことは？
     - 内容の簡単なサマリ
     - 重要な公式のまとめ など
   - 振り返りの結果は紙やテキストファイルにまとめましょう

Day 3

確率分布

目次

第１章：確率変数と確率分布
第２章：離散型確率分布
第３章：連続型確率分布
第４章：修了演習
- ロジスティック回帰

確率変数と確率分布について学ぼう

代表的な確率分布を学ぼう

本日学んだ内容を
機械学習に応用しよう！

第１章

確率変数と確率分布

確率変数

確率変数：確率的に値が決まる変数

例) フェアなコインを投げて、表ならば１、裏ならば０となる変数 X を考える

$$P(X = 0) = \frac{1}{2}$$

$$P(X = 1) = \frac{1}{2}$$

X の値は確率的に決まる

X は確率変数

確率変数は２種類ある

離散変数

- 取りうる値が「とびとび」な確率変数
- 例) 植物の種子の数

連続変数

- 取りうる値が「なめらか」な確率変数
- 例) 身長

確率分布

確率分布：確率変数のとりうる値とその値の出現確率の対応関係

以下の２つが定まっていることが大事

1. 確率変数があり、取りうる値が定まっている

$$X = 1, 2, 3, 4 \ldots$$

2. それらの値の出現確率がそれぞれ決まっている

$$P(X = 1) = \frac{1}{6}, P(X = 2) = \frac{1}{6}, \ldots$$

確率分布

例）コインを２枚投げ、表が出た枚数を X とする

Xの取りうる値：　？

出現確率：$P(X=0)=$

？

$P(X=1)=$

$P(X=2)=$

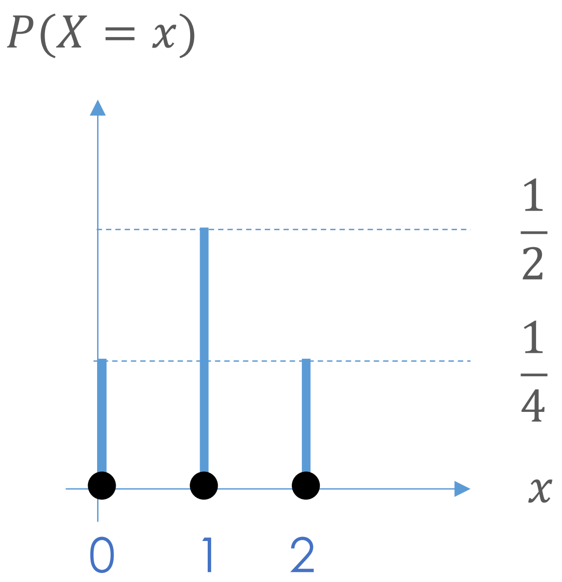
例）コインを２枚投げ、表が出た枚数を X とする

Xの取りうる値：$0, 1, 2$

出現確率： $P(X=0)=\frac{1}{4}$

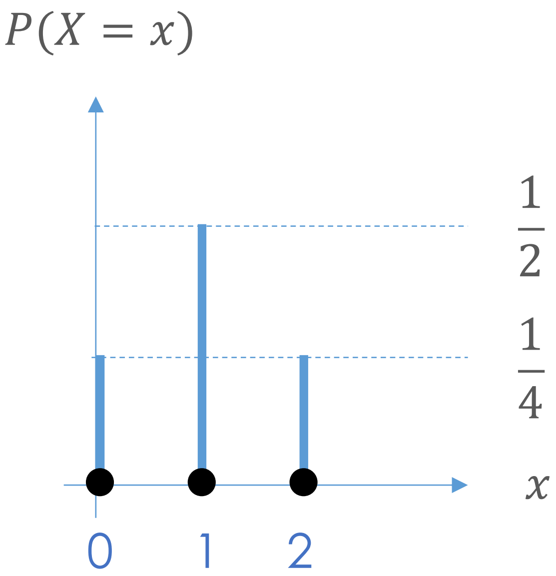
$P(X=1)=\frac{1}{2}$

$P(X=2)=\frac{1}{4}$



確率分布

コインを２枚投げた場合における表が出た枚数 X の分布を求めました



Xの出方は無秩序ではなく

確率分布 $P(X)$ によって制御されている

- $X = 1$ となりやすい
- Xは 3 や 4 にはならない

このような状況を「X は $P(X)$ に従う」と表現する

確率分布

データ X はどのような確率分布に従っているだろうか？推測してみよう！

$$n = 1200$$

X	出現頻度
1	160
2	400
3	160
4	160
5	160
6	160

確率分布

データ X はどのような確率分布に従っているだろうか？推測してみよう！

$$n = 1200$$

X	出現頻度
1	160
2	400
3	160
4	160
5	160
6	160

$$P(X = 2) = \frac{1}{3}$$

$$P(X = i) = \frac{2}{15}\ (i = 1, 3, 4, 5, 6\ )$$

2 が出やすいイカサマ６面サイコロ？

データが従う確率分布を求め予測や意思決定に応用することは、統計・機械学習の活用方法の一つ

確率変数の代表値

確率変数の代表値

- 確率変数 X の動きは、指標を使って要約することも可能
- 手元にあるデータを要約するときと、同じ考え方
  「中心」と「ばらつき」を考える

期待値：$E[X] = \mu = \sum x\, p(x)$

分散：$V[X] = \sigma^2 = \sum (x - \mu)^2\, p(x)$

標準偏差：$\sigma = \sqrt{\sigma^2}$

確率変数の代表値

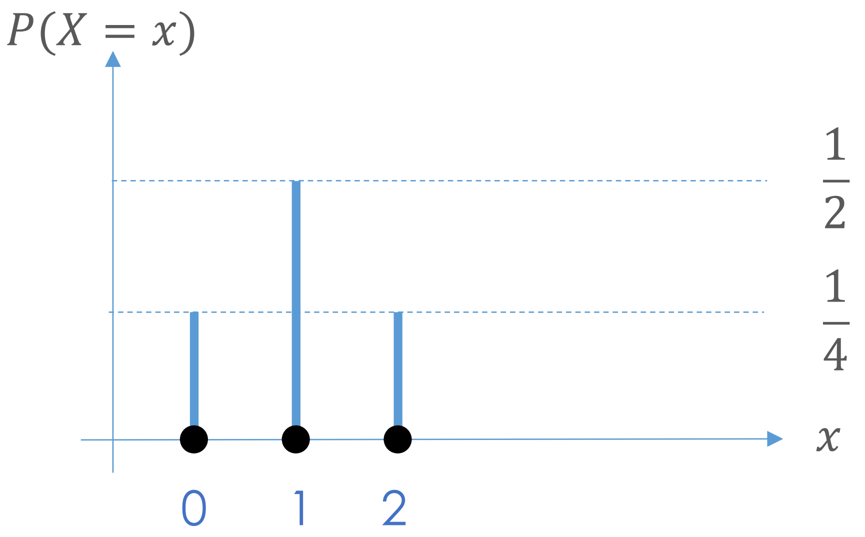
<u>確率変数の期待値</u>

$$E[X] = \sum_x xP(X = x)$$

例）

$$P(X = 0) = \frac{1}{4}, P(X = 1) = \frac{1}{2}, P(X = 2) = \frac{1}{4}$$

$$E[X] = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$$



確率変数の代表値

分散：期待値からどれほどばらつくかを示す指標

分散 $$V[X] = \sum_{x} (x - E[X])^2 P(X = x)$$

標準偏差 $$\sqrt{V[X]} = \sqrt{\sum_{x} (x - E[X])^2 P(X = x)}$$

確率変数の代表値

x	$P(X = x)$
0	$\frac{1}{10}$
1	$\frac{1}{5}$
2	$\frac{2}{5}$
3	$\frac{1}{5}$
4	$\frac{1}{10}$

$$E[X] = 2$$

$$V[X] = 1.2$$

$$\sqrt{V[X]} = \sqrt{1.2} \approx 1.1 \ldots$$

期待値 2 から 1.1 程度値がばらつく

確率変数の代表値

連続型の確率変数の場合にはシグマが積分に変わる！

$$E[X] = \int x\, P(X = x)dx$$

$$V[X] = \int (x - E[X])^2 P(X = x)dx$$

第１章：理解確認

（１）確率分布を自作せよ（確率分布と呼ぶためには何が必要であったか考えよ）

（２）以下のようなくじを１本引いたときの賞金を X(円) とする

	本数	賞金
１等	1	１２００円
２等	2	６００円
３等	3	９０円

（はずれは無しとする）

(2-1) $E[X]$ を求めよ

(2-2) $V[X]$ を求めよ

(2-3) 賞金額を全て300円増加させた

$E[X]$ と $V[X]$ を計算せずに求めよ

（１）「取りうる値」と「その値の出現確率」が定まっていれば確率分布である

例）$X = 1,2,3,4$　　$P(X = x) = \frac{1}{4}$　$(x = 1,2,3,4)$　　公平な４面サイコロ

（２-１）　$P(X = 1200) = \frac{1}{6}, P(X = 600) = \frac{1}{3}, P(X = 90) = \frac{1}{2}$

$$E[X] = 1200 \times \frac{1}{6} + 600 \times \frac{1}{3} + 90 \times \frac{1}{2} = 445$$

（２-２）　$V[X] = (1200 - 445)^2 \times \frac{1}{6} + (600 - 445)^2 \times \frac{1}{3} + (90 - 445)^2 \times \frac{1}{2} = 166025$

（２-３）　$E[X] = 745$

$V[X] = 166025$

確率変数の取りうる値が全て一律増えるので

平均値は底上げされても、ばらつき度合いは変わらない

第２章
離散型確率分布

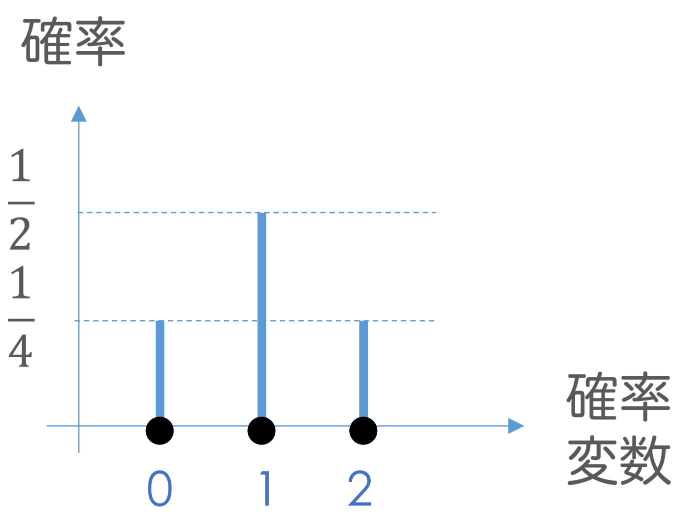
確率分布

実は、これまで見てきた確率分布は「離散型の確率変数」に対する確率分布

確率分布

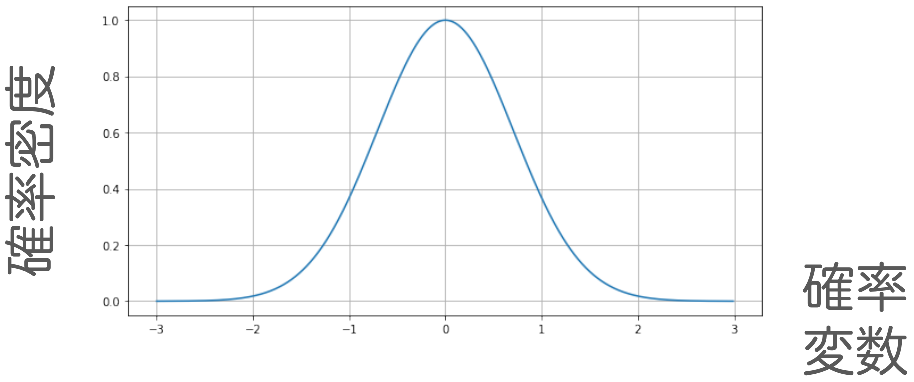
離散型の確率分布

(確率質量関数とも呼ぶ)



連続型確率分布

(確率密度関数とも呼ぶ)



離散型確率分布

まず最初に抑えるべき離散型確率分布

- 離散一様分布
- ベルヌーイ分布
- 二項分布
- ポアソン分布

使用用途の例：ポアソン回帰

目的変数がポアソン分布に従うと仮定し、目的変数と説明変数の期待値を式で結ぶ

離散一様分布

離散一様分布

$$P(X = x) = \frac{1}{n} \; (x = 1, \dots, n)$$

全ての事象が等確率で出現

$n = 6$ でフェアなサイコロ投げになる

ベルヌーイ分布：コイン投げの一般化

これまでのコイン投げ

- コインの表が出る確率は $\frac{1}{2}$
- コインの裏が出る確率は $\frac{1}{2}$

一般化されたコイン投げ

- コインの表が出る確率は $p\ (0 \leq\ p \leq\ 1)$
- コインの裏が出る確率は $q\ (q = 1 - p)$

ベルヌーイ分布

「表が出る」ことを 1 、「裏が出る」ことを 0 とすると

$P(X=1)=p,\ P(X=0)=q=1-p$

1 行でまとめると

ベルヌーイ分布

$$P(X=x)=p^{x}(1-p)^{1-x}\ (x=0,1)$$

ベルヌーイ分布

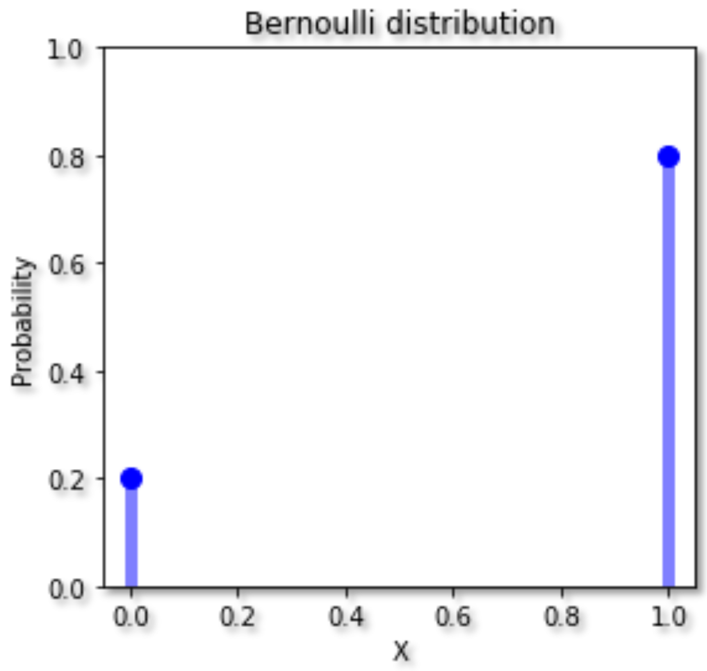
$$P(X = 1) = p = 0.8$$

ベルヌーイの分布の期待値・分散・標準偏差

平均：$\mu = p$

分散：$\sigma^2 = p(1 - p)$

標準偏差：$\sigma = \sqrt{p(1 - p)}$



ベルヌーイ分布の発展

公平なコインを６回投げた

Y_1 Y_2 Y_3 Y_4 Y_5 Y_6

表が出れば $Y = 1$

裏が出れば $Y = 0$

表が出た総数を X とする

X はどのようにして計算できる？

ベルヌーイ分布の発展

公平なコインを６回投げた

Y_1 Y_2 Y_3 Y_4 Y_5 Y_6

表が出れば $Y=1$

裏が出れば $Y=0$

表が出た総数を X とする

X はどのようにして計算できる？

$$X = Y_1 + Y_2 + \cdots + Y_6$$

ベルヌーイ分布の発展

X はどのような範囲の値をとる？

最も起こりやすい X の値は？

最も起こりにくい X の値は？

$X = 2$ となる確率は？

ベルヌーイ分布の発展

X はどのような範囲の値をとる？　$X = 0, 1, 2, \dots, 6$

最も起こりやすい X の値は？　$X = 3$

最も起こりにくい X の値は？　$X = 0 \text{ or } 6$

$X = 2$ となる確率は？　${}_6C_2 \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{6-2} \approx 0.23$

確率変数 X は何かしらの分布により制御されている

この X が従う確率分布が二項分布

二項分布

二項分布

$$P(X = x) = {}_nC_x \, p^x (1-p)^{n-x} (x = 0, 1 \ldots, n)$$

- ベルヌーイ分布に従う確率変数 $Y_1, Y_2, \cdots, Y_n$ の「和」が従う分布
- $Y_1, Y_2, \cdots, Y_n$ は全て同じベルヌーイ分布に従うという前提（同じコインをふる）

二項分布

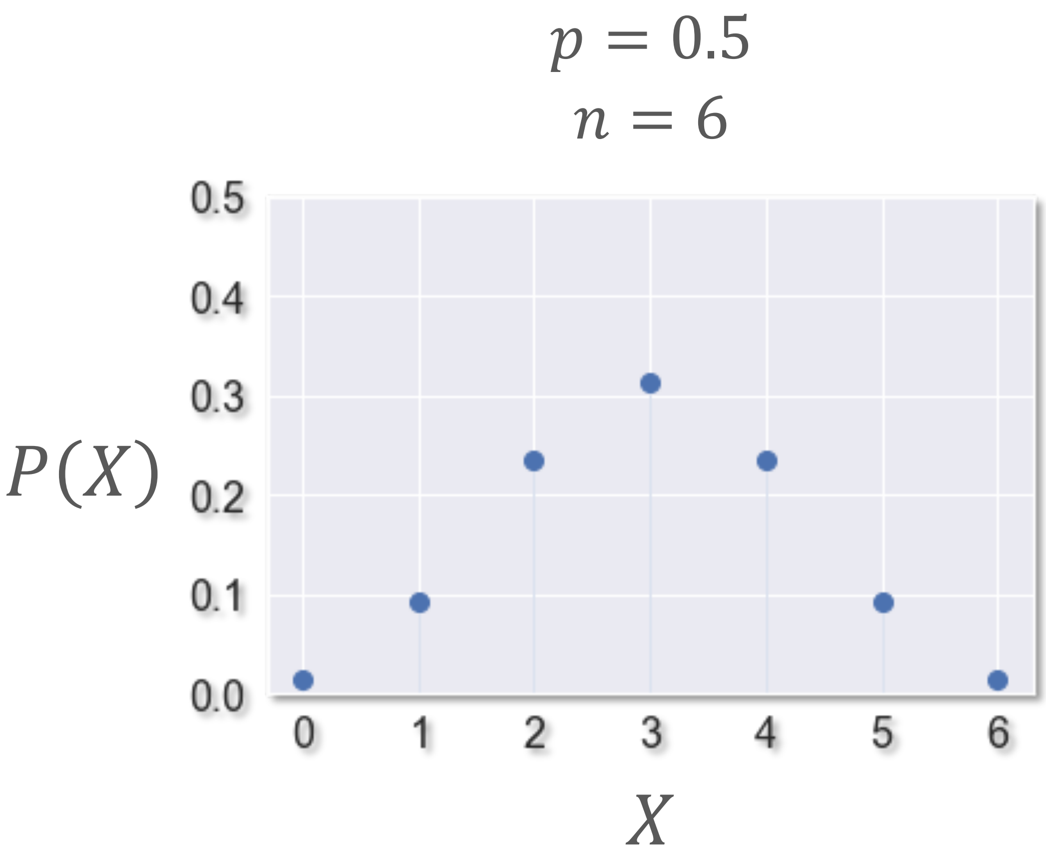
$$P(X=x)={}_nC_x\,p^x(1-p)^{n-x}\,(x=0,1\ldots,n)$$

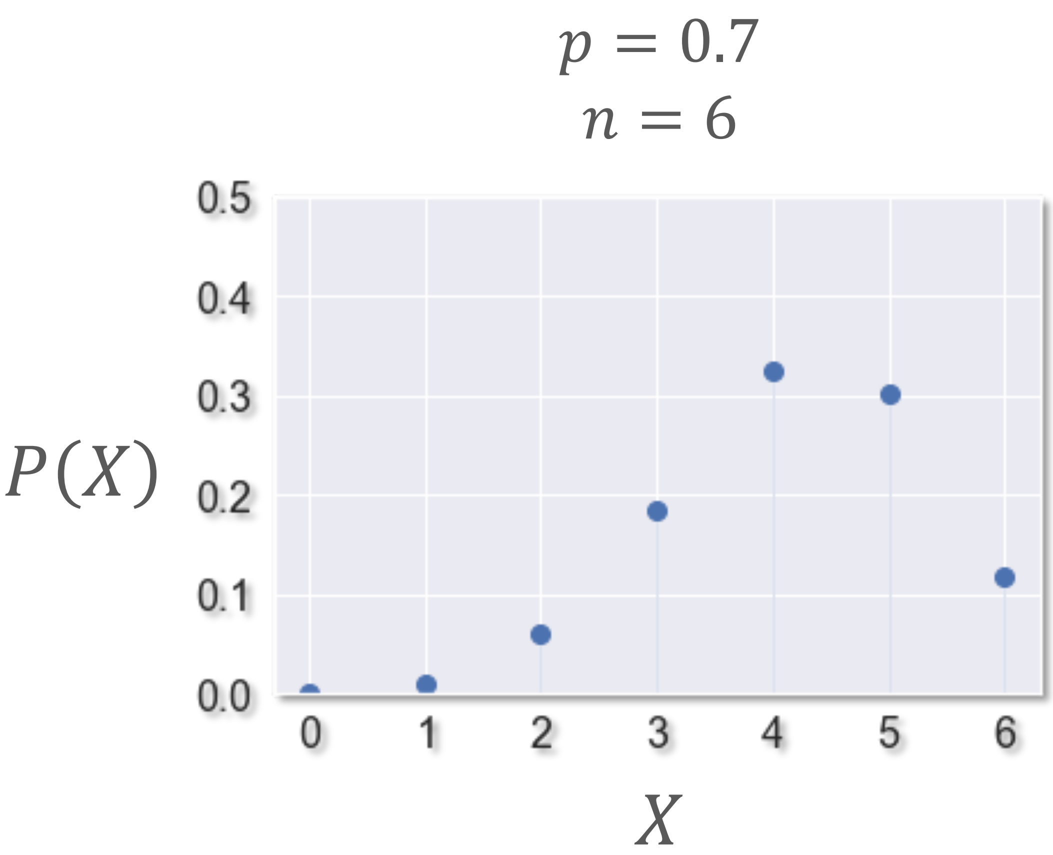
${}_nC_x$：異なる n 個のものから x 個を選ぶ組み合わせの総数

p^x：表が x 回出る確率

$(1-p)^{n-x}$：裏が $(n-x)$ 回出る確率

二項分布


$p = 0.5$
$n = 6$
0.5
0.4
0.3
0.2
0.1
0.0
$P(X)$
0 1 2 3 4 5 6
X


$p = 0.7$
$n = 6$
0.5
0.4
0.3
0.2
0.1
0.0
$P(X)$
0 1 2 3 4 5 6
X

二項分布の発展

二項分布

- 確率変数 X の上限/下限が定まっている確率的現象を表現
- 例）コインを n 回投げると x 回表が出た

お店の来店者数　　これは二項分布で表現できるだろうか？

二項分布の発展

二項分布

- 確率変数 X の上限/下限が定まっている確率的現象を表現
- 例）コインを n 回投げると x 回表が出た

お店の来店者数　　これは二項分布で表現できるだろうか？

上限はなさそう？

そのような現象を表現する分布はポアソン分布

ポアソン分布

ポアソン分布

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \ (x = 0, 1, \dots)$$

- λ は X の平均値（期待値）を定めるパラメータ
- X の取りうる値に上限はない
- この式は（ある制約の下で）二項分布の n を無限大にすると理論的に導かれる

ポアソン分布

ポアソン分布の意味：

ある期間内に平均 λ 回起こる事象において、その期間に X 回起こる確率の分布

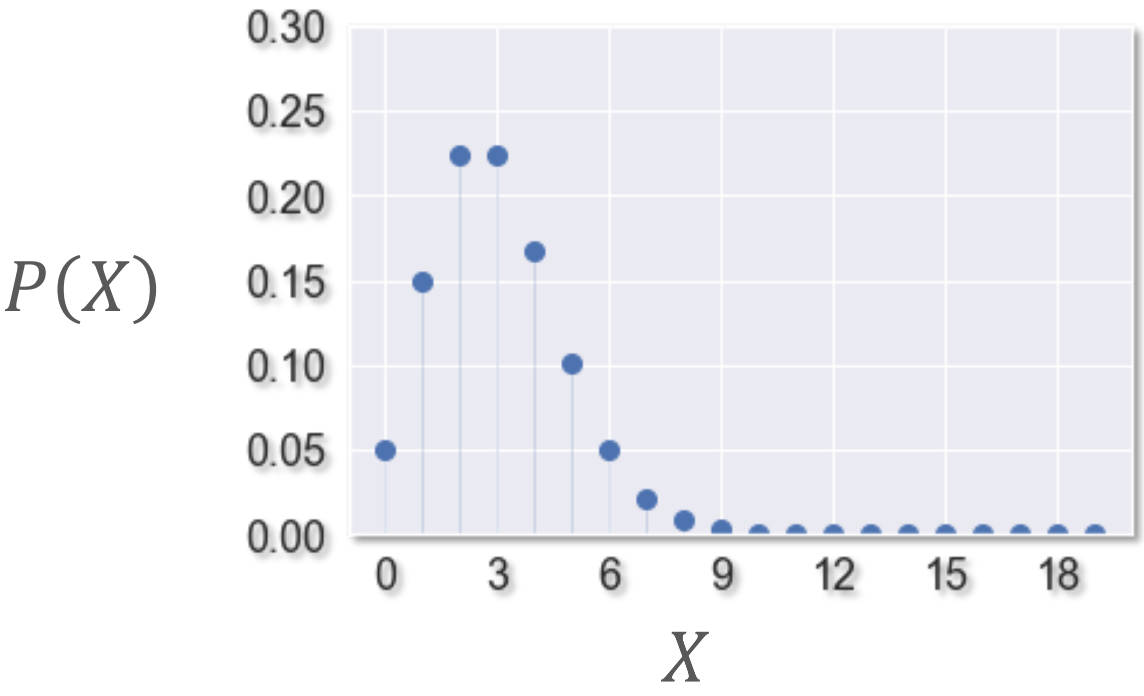
１時間に平均３人の来店があるお店について考える

このお店に１時間に x 人来店する確率は $\frac{3^x}{x!}e^{-3}$ である

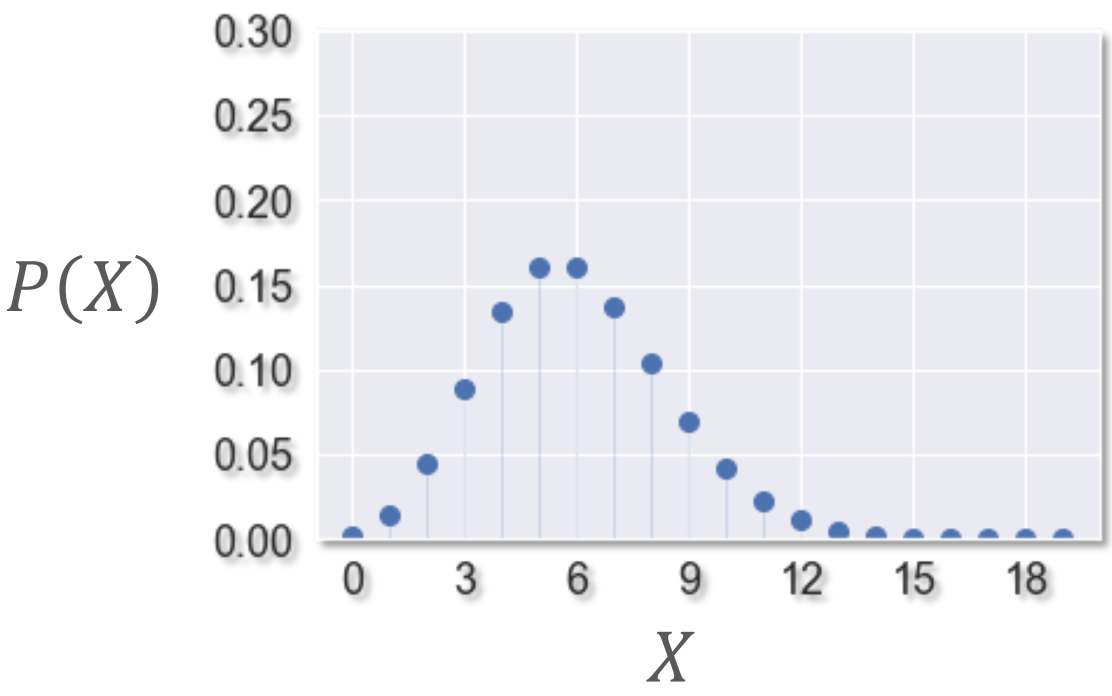
１時間に１人も来店しない確率は $\frac{3^0}{0!}e^{-3} \approx 0.0498$である

ポアソン分布

$\lambda = 3$



$\lambda = 6$



第２章：理解確認

（１）　確率変数 Y_1 と Y_2 は、$p = 0.2$ のベルヌーイ分布に従う

$X = Y_1 + Y_2$ が従う確率分布を求めよ

また $X = 1$ となる確率を求めよ

（２）　ある道路で交通量調査を3時間行い、トラックが30台通ったとする

トラックの通過台数 X がポアソン分布に従うと仮定する

(2-1) トラックの平均通過率 λ (台/時) を求めよ

(2-2) その道路において１時間でトラックが5台通る確率を求めよ

（３） 以下の実験結果が得られる確率を求めよ

とある実験の結果

試行	1	2	3	4	5	6	7	8	9	10
x	1	1	1	0	1	1	1	1	0	1

- $P(X=x)=p^x(1-p)^{1-x}$
- $X=1$ となる確率を p とする
- x_i を i 回目の試行における X の値とする
- 各試行は独立とする

（１） X は $p = 0.2, n = 2$ の二項分布に従う

$$P(X = 1) = {}_2C_1\,(0.2)^1(1 - 0.2)^{2-1} = 0.32$$

（２-１）平均通過率 λ (台/時) $= \frac{30}{3} = 10$

（２-２）$P(X = 5) = \frac{10^5}{5!}e^{-10} \approx 0.037$

（３） $\prod_{i=1}^{10} p^{x_i}(1 - p)^{1-x_i} = p^8(1 - p)^2$

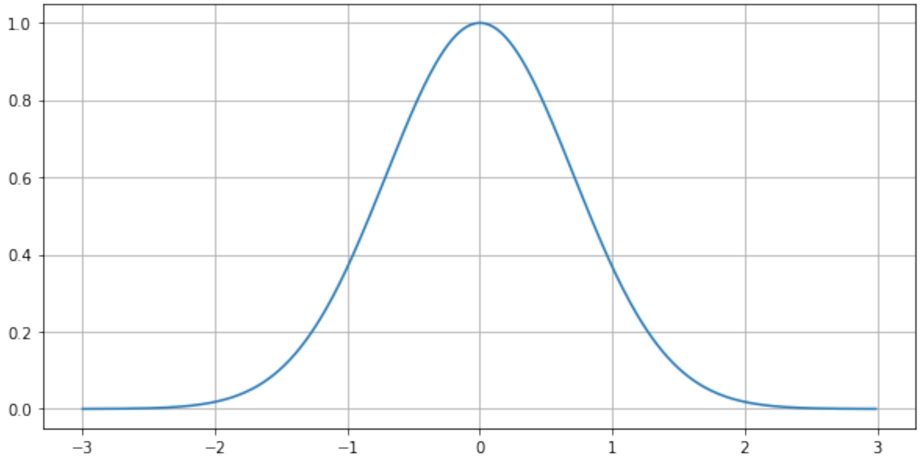
第３章

連続型確率分布

連続型確率分布

- 確率変数が連続変数（なめらか）である分布
- 確率の代わりに「確率密度」という考え方を用いることに注意！

確率密度 $p(x)$

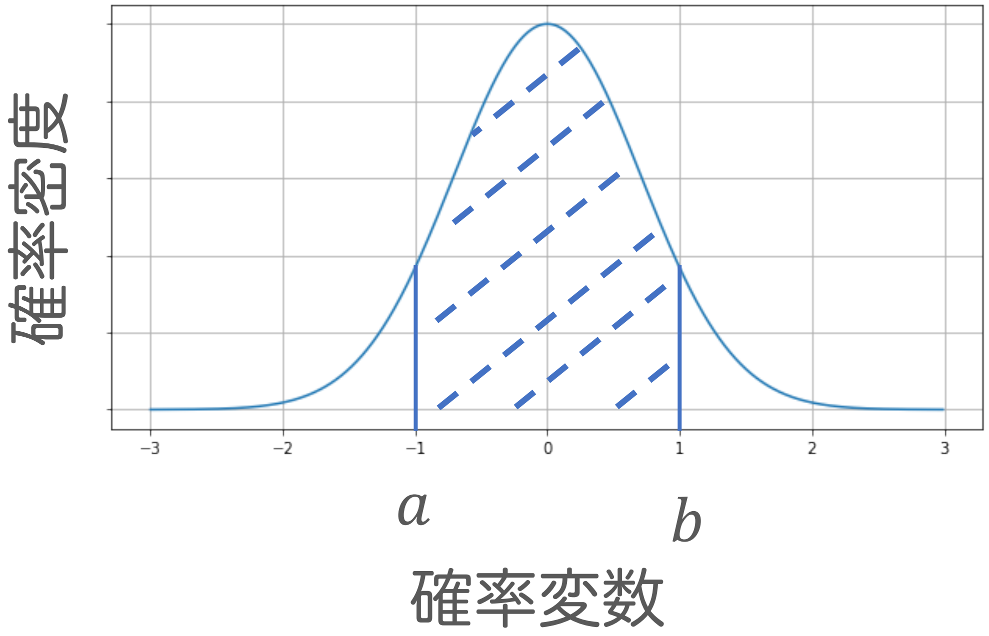


確率変数 x

連続型確率分布

確率密度

- 対応する確率変数の値の相対的な出やすさ
- 確率密度が大きい確率変数の値の方が他と比べて出やすい



- 確率変数の値が a 以上 b 以下となる確率が斜線領域の面積の値で求まる
  - 確率変数が取りうる全区間の面積を計算すると1になるようにできている
- それゆえ縦軸は確率密度という呼称になる

連続型確率分布

確率ではだめな理由のイメージ

０～９の整数が等確率で入る箱がある

1	2	3

となる確率は？

連続型確率分布

確率ではだめな理由のイメージ

0～9の整数が等確率で入る箱がある

1	2	3

となる確率は？ $\dfrac{1}{10^3}$

連続値とは無限長の箱に入った数字

$1 = \boxed{1.}\boxed{0}\cdots\boxed{0}$ $\qquad \dfrac{1}{10^\infty} \approx 0$

連続変数に対して確率を割り当てても全て０となり無意味

連続型確率分布

まず最初に抑えるべき連続型確率分布

- 連続一様分布
- 正規分布
- 指数分布

特に正規分布は必ず理解しましょう

連続一様分布

連続一様分布

連続一様分布

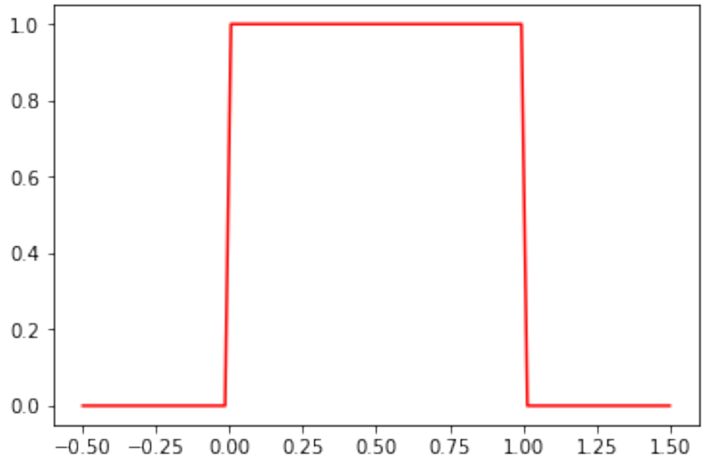
$$p(x) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ 0 & (\text{その他}) \end{cases} \qquad a, b\text{は定数}$$

どんなグラフになるかイメージつきますか？

連続一様分布

$a = 0,\ b = 1$の場合

$p(x)$



x

$a = 0$ $b = 1$

$$p(x) = \begin{cases} 1 & (a \leq x \leq b) \\ 0 & (\text{その他}) \end{cases}$$

連続一様分布は「ある範囲の値が出やすいとも出にくいとも言っていない」

という見方ができ、確率分布に対して全く知識がない時の仮定として使える

指数分布

指数分布

$$p(x) = \lambda e^{-\lambda x} \ (x \geq 0)$$

- 確率的に発生する事象の「**発生時間間隔**」を表す分布
- λ は単位時間における事象の発生回数を表すパラメータ

どんな現象を表現するために使えそう？

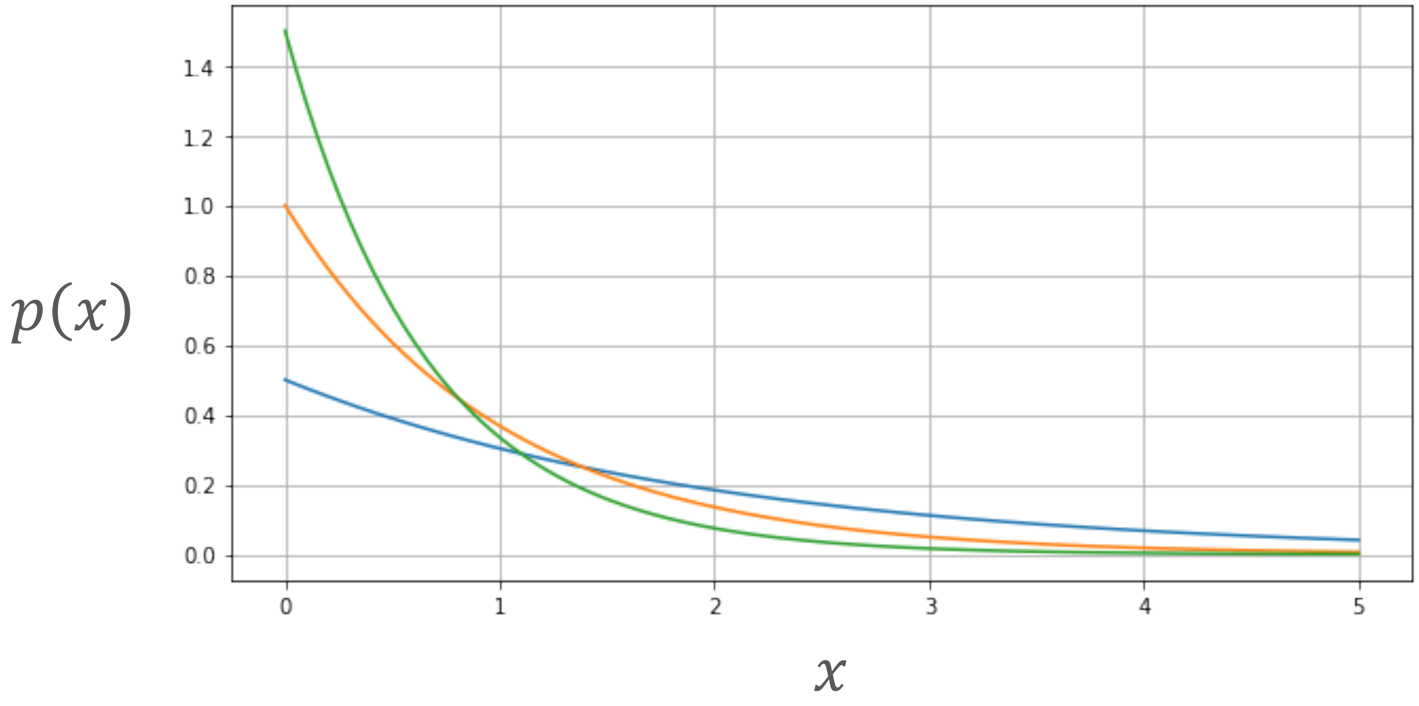
指数分布

指数分布で表現できる事象の例）

- 地震が起きる間隔
- メールを受信するタイミングの間隔
- お店にお客さんが来るタイミング

指数分布

$$p(x) = \lambda e^{-\lambda x} \ (x \geq 0)$$



$\lambda = 0.5, \quad \frac{1}{\lambda} = 2$

$\lambda = 1.0, \quad \frac{1}{\lambda} = 1$

$\lambda = 1.5, \quad \frac{1}{\lambda} = \frac{2}{3}$

正規分布

正規分布

正規分布（ガウス分布）

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

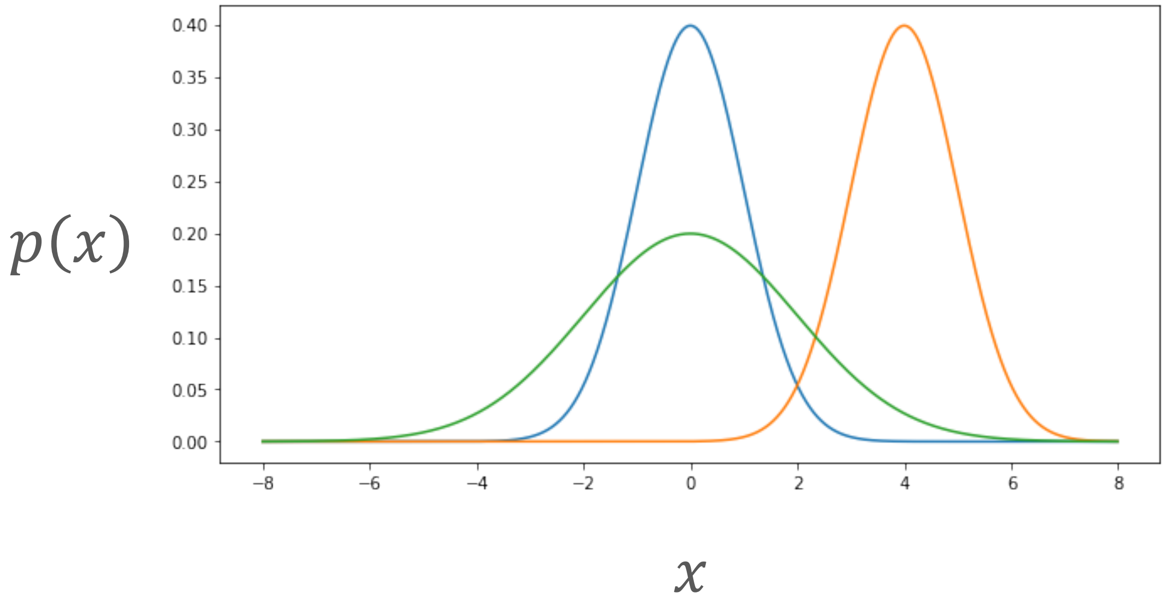
μ：平均

σ^2：分散

正規分布

山なり形状がポイント！

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



$\mu = 0, \sigma^2 = 1$

$\mu = 4, \sigma^2 = 1$

$\mu = 0, \sigma^2 = 4$

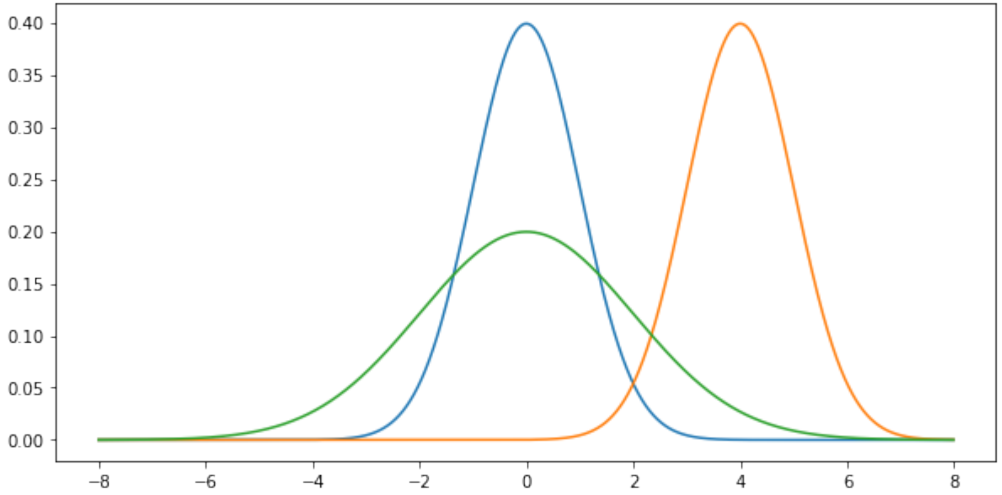
どの分布に対応している？

正規分布

山なり形状がポイント！

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

$p(x)$



x

$\mu = 0, \sigma^2 = 1$

$\mu = 4, \sigma^2 = 1$

$\mu = 0, \sigma^2 = 4$

正規分布

「山なり形状」はどこから出てくる？

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

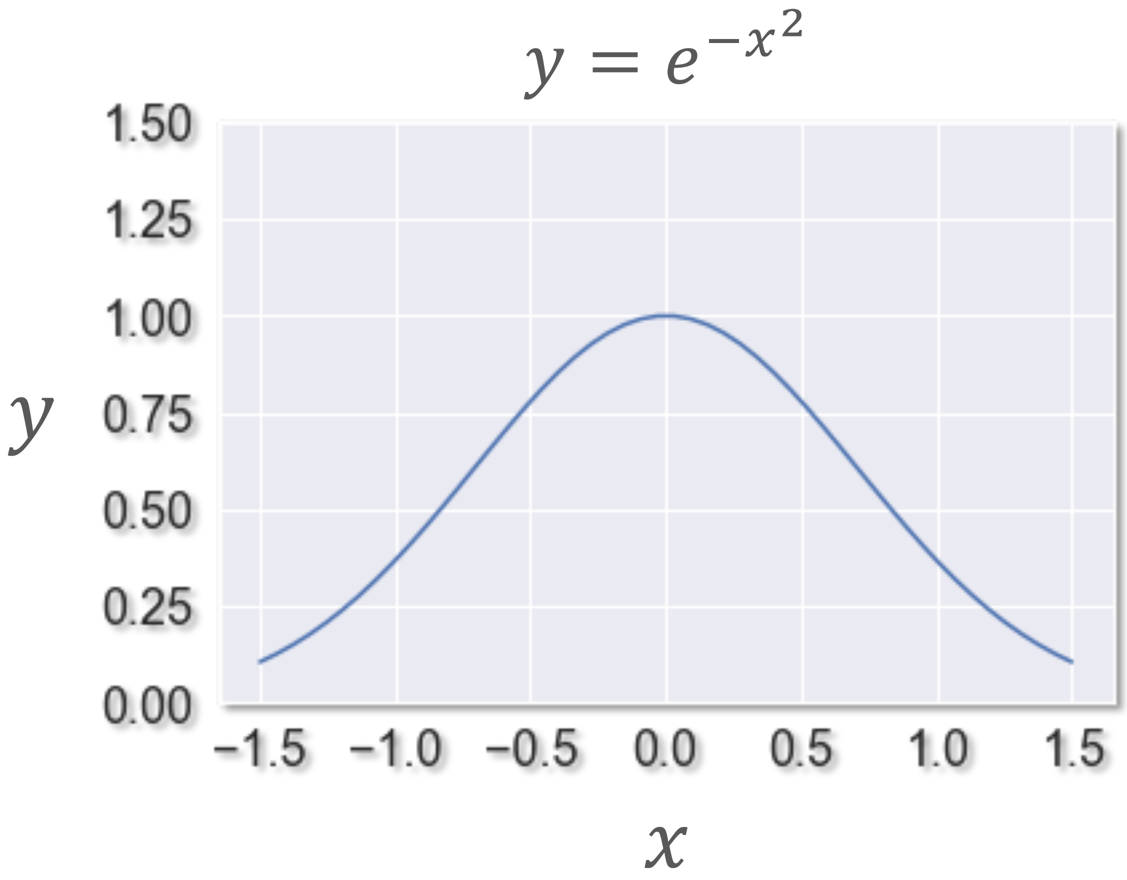
正規分布

「山なり形状」はどこから出てくる？

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

ここが山をつくる！

正規分布



山の頂点と開き具合を変えられるように、 $-x^2$ にー工夫したのが $-\frac{(x-\mu)^2}{2\sigma^2}$

正規分布

$e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ でも良いのか？

正規分布

$e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ でも良いのか？　確率分布としてはNG

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi\sigma^2}$$ となる

確率の公理である「すべて足すと１」を満たさない

正規分布

$e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ でも良いのか？　確率分布としてはNG

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi\sigma^2}$$ となる

確率の公理である「すべて足すと１」を満たさない

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

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \sqrt{2\pi\sigma^2} = 1$$

正規分布

そもそもなぜ「山なり形状」が必要なのか？

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

正規分布

そもそもなぜ「山なり形状」が必要なのか？

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

正規分布が表したい現象

- 平均付近の値が出やすい
- 平均から離れるにつれて相対的な出やすさが左右対称に小さくなる

確かにこのような出方をする確率変数は多そう

正規分布の性質

正規分布については以下の内容も理解しておこう

- 「データが正規分布に従う」と仮定すると解析解を得やすい
  （コンピューターに頼らず手計算で答えが出る）
- σ 区間と呼ばれる区間を考えると
  確率変数の値がある範囲に収まる確率がざっくり分かる

正規分布の性質

- 解析解：実際に手計算をすることで得られる厳密な解
  - 例）６面のサイコロを１回振った時、１が出る確率は？　→ 1/6（解析解）
  - 確率分布に正規分布を仮定すると解析解が得られること多々ある

- 数値解：コンピュータを用いたシミュレーションによって得られる解
  - 解に誤差が含まれている
  - 例）1. 気体分子の運動のシミュレーション
    2. 複雑な確率分布に従う確率変数の期待値を求める

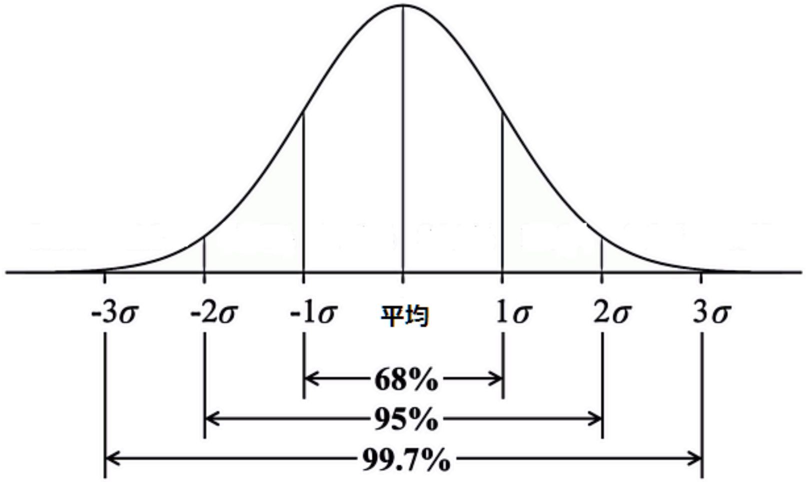
正規分布の性質

正規分布は標準偏差 σ に対して以下のような性質を持つ

平均 μ 標準偏差 σ の正規分布に従う確率変数の値が

- 平均 $\pm\,\sigma$の区間に入る確率は約68%
- 平均 $\pm\,2\sigma$の区間に入る確率は約95%
- 平均 $\pm\,3\sigma$の区間に入る確率は約99.7%

である



（１）　連続型確率分布に従う確率変数の場合、値が a 以上 b 以下 となる確率は

$$\int_a^b p(x)dx$$

で与えられる

$p(x)$ が指数分布の場合

$$\int_a^b p(x)dx = \int_a^b \lambda e^{-\lambda x}dx = -e^{-\lambda b} + e^{-\lambda a}$$

となる

これを用いて、$\lambda = 1$の指数分布に従う確率変数 X が 2 以上 3 以下となる確率を求めよ

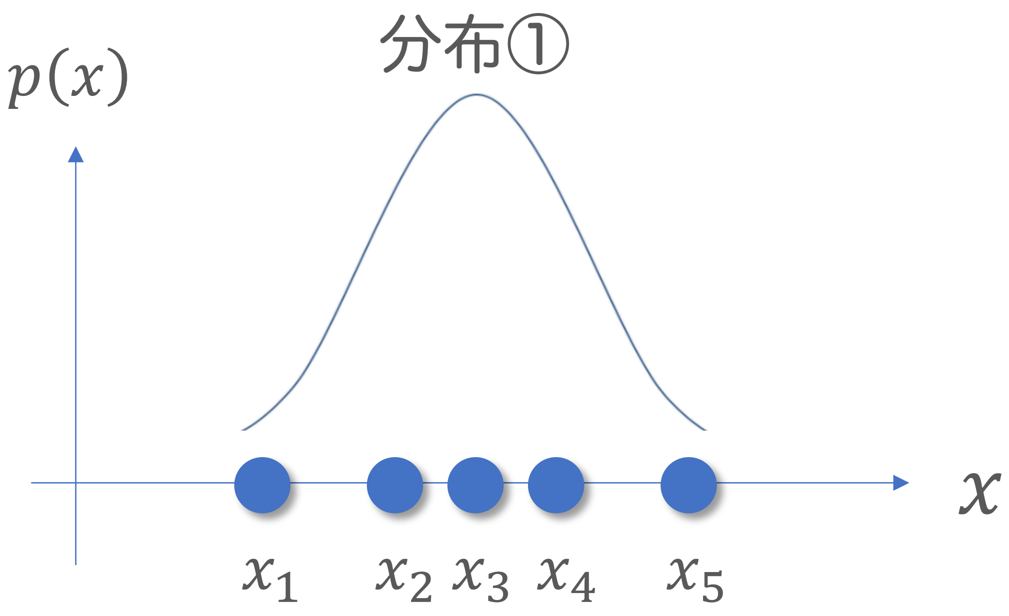
以下の正規分布 $p(x)$ について次の問いに答えよ

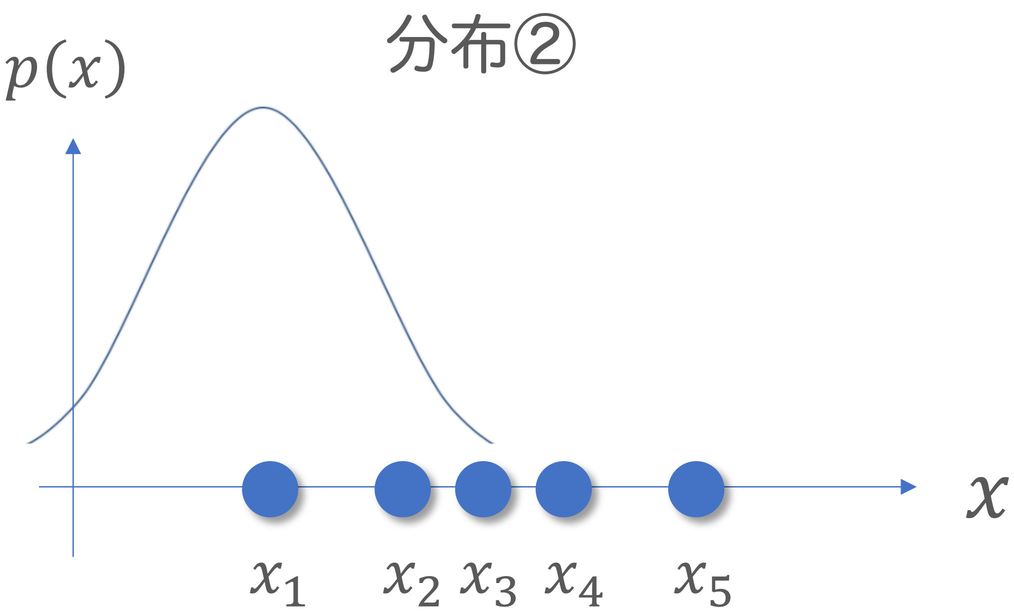
$$p(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{(x-10)^2}{8}}$$

（２）平均 μ と標準偏差 σ を求めよ

（３）正規分布 $p(x)$ に従う確率変数 X の値が 10 以上となる確率を求めよ
ヒント：正規分布の形状を元に考えよ（積分は必要ない）

（４） x_1, x_2, x_3, x_4, x_5 は正規分布 $p(x)$ に従うとする
５つのデータから分布の様子を見積もりたい
分布②よりも分布①の方が尤もらしい分布に思えるが
その尤もらしさをどのように定量化すれば良いだろうか





第３章：理解確認

（１）　$\int_2^3 e^{-x}dx = -e^{-3} + e^{-2} \approx 0.086$

（２）　$p(x) = \frac{1}{\sqrt{2\pi 2^2}} e^{-\frac{(x-10)^2}{2 \cdot 2^2}}$　より $\mu = 10,\ \sigma = 2$

（３）　$p(x \geq 10) = \frac{1}{2}$

正規分布 $p(x)$ は平均値 10 を中心に左右対称であるため

（４）　$\prod_{i=1}^{5} p(x_i)$ を計算すると尤もらしい分布（分布①）の方が大きな値となる

確率密度関数の「高さ」を利用して尤もらしさを測っている

第４章
修了演習

修了演習

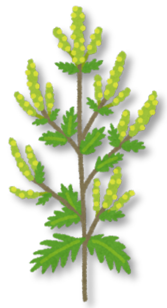
本日学んだ内容を機械学習に応用しよう！

- ロジスティック回帰

※ 付録に正規分布を用いた異常検知の問題があります
　ぜひ挑戦してみましょう

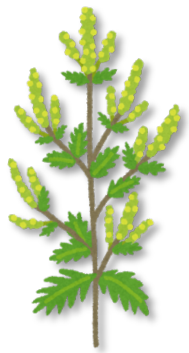
ロジスティック回帰

ロジスティック回帰



とある植物は、身体のサイズが大きいほど
種子を持つ割合が高いように思える

i 番目の植物



x_i ：身体のサイズ

$t_i = 1$ ：種子を持つ

$t_i = 0$ ：種子を持たない

$T = t_i \ (t_i = 0 \text{ or } 1)$ として

$P(T)$ が従う確率分布を求めたい

ロジスティック回帰

$t_i = 1$：種子を持つ

$t_i = 0$：種子を持たない

２値の現象を表現するのが得意な確率分布があった

ロジスティック回帰

$t_i = 1$：種子を持つ

$t_i = 0$：種子を持たない

２値の現象を表現するのが得意な確率分布があった

ベルヌーイ分布

$$P(X = x) = p^x(1 - p)^{1-x} \ (x = 0, 1)$$

ロジスティック回帰

$$P(T = t_i) = y^{t_i}(1 - y)^{1-t_i} \ (t_i = 0, 1)$$

y は $t_i = 1$ となる確率

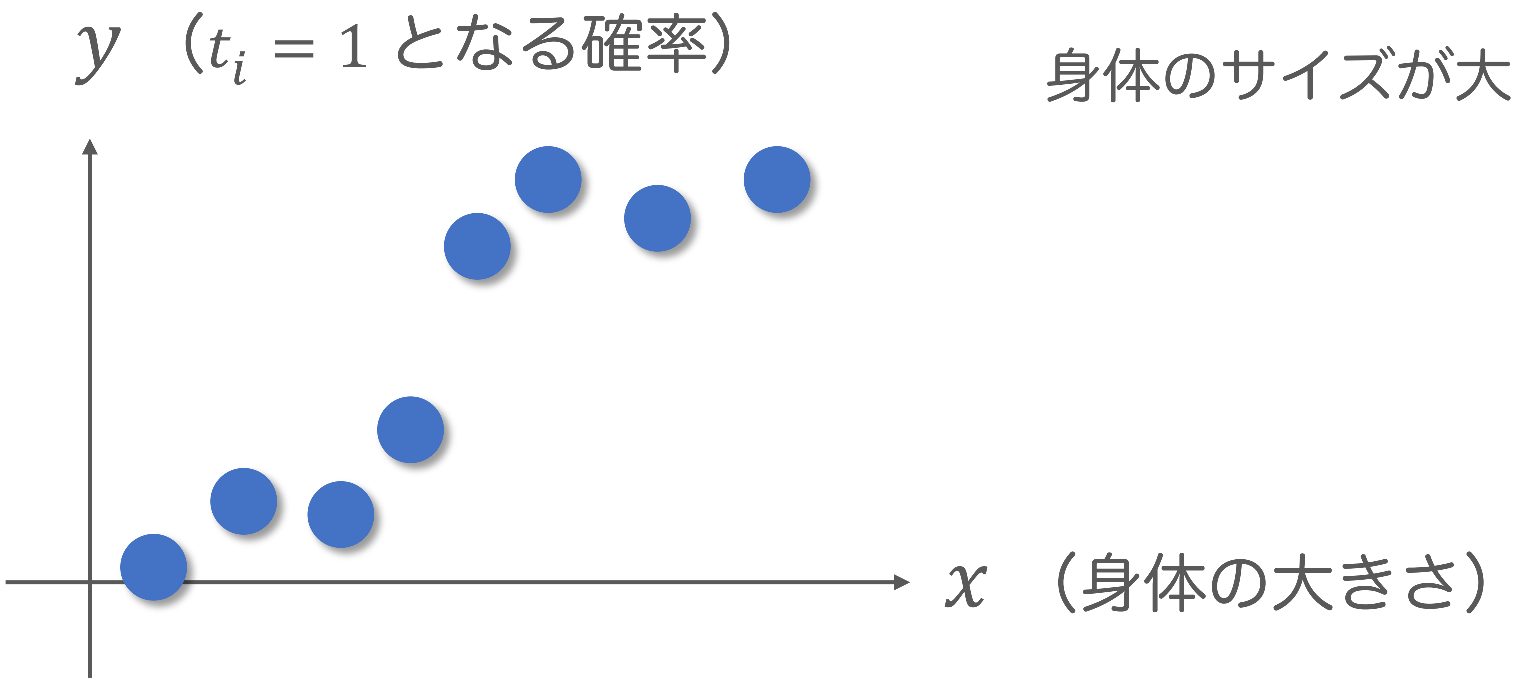
y が身体の大きさ x_i に依存するなら

$y_i = f(x_i)$ のようにすれば良い？

$$P(T = t_i | x_i) = y_i^{t_i}(1 - y_i)^{1-t_i} \quad y_i = f(x_i)$$

ロジスティック回帰

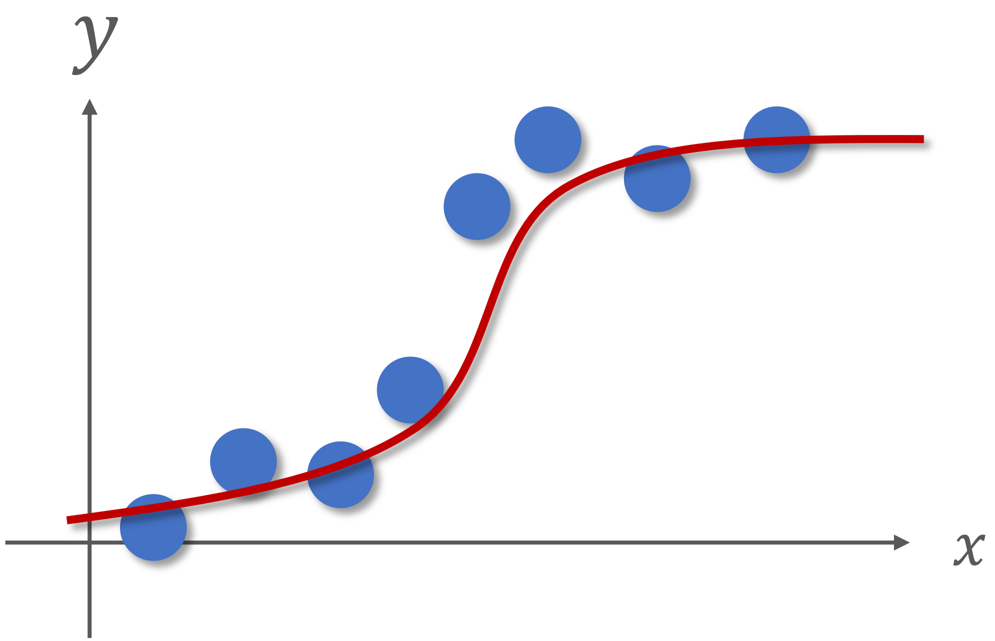
関数 f をどのように設定すべきか？



身体のサイズが大きい程種子を持つ割合は高い

ロジスティック回帰

関数 f をどのように設定すべきか？

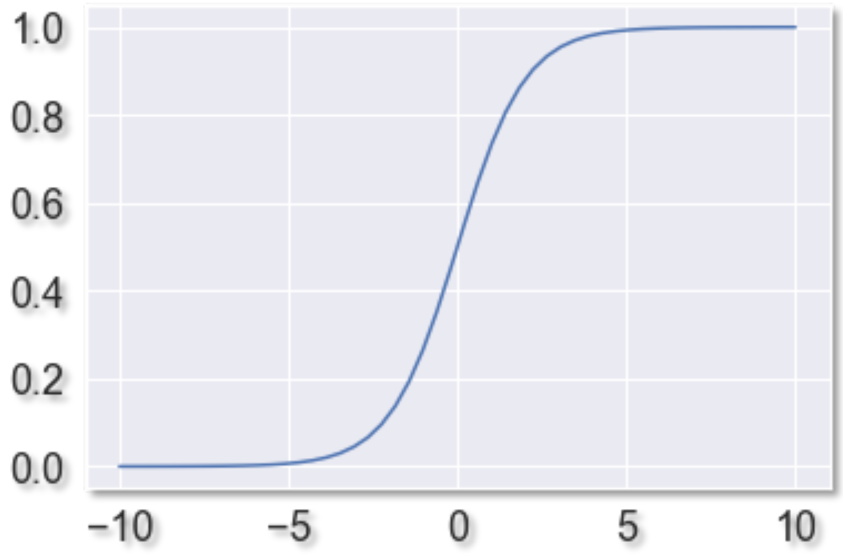


身体のサイズが大きい程種子を持つ割合は高い

シグモイド関数を用いる

ロジスティック回帰

$$y = \frac{1}{1 + e^{-wx}}$$



シグモイド関数

- 出力が０～１の連続値
- 出力を確率値として解釈できる
- w によって曲がり具合を制御できる

ロジスティック回帰

ロジスティック回帰

$$P(T = t_i | x_i) = y_i^{t_i} (1 - y_i)^{1 - t_i} \quad y_i = \frac{1}{1 + e^{-w x_i}}$$

T は 入力 x を受け付ける不思議議なコイン

x の値によって $t = 1$ となる確率 y が変わる

y と x の関係はシグモイド関数で結ばれている

ロジスティック回帰

入力 x_i を受けて t を表現する確率モデル $P(T = t_i|x_i)$ が定まった

データを用いてパラメータ（学習すべきもの）を求めれば良い

何がパラメータであろうか？

$$P(T = t_i|x_i) = {y_i}^{t_i}(1 - y_i)^{1-t_i} \quad y_i = \frac{1}{1 + e^{-wx_i}}$$

ロジスティック回帰

入力 x_i を受けて t を表現する確率モデル $P(T = t_i|x_i)$ が定まった

データを用いてパラメータ（学習すべきもの）を求めれば良い

何がパラメータであろうか？

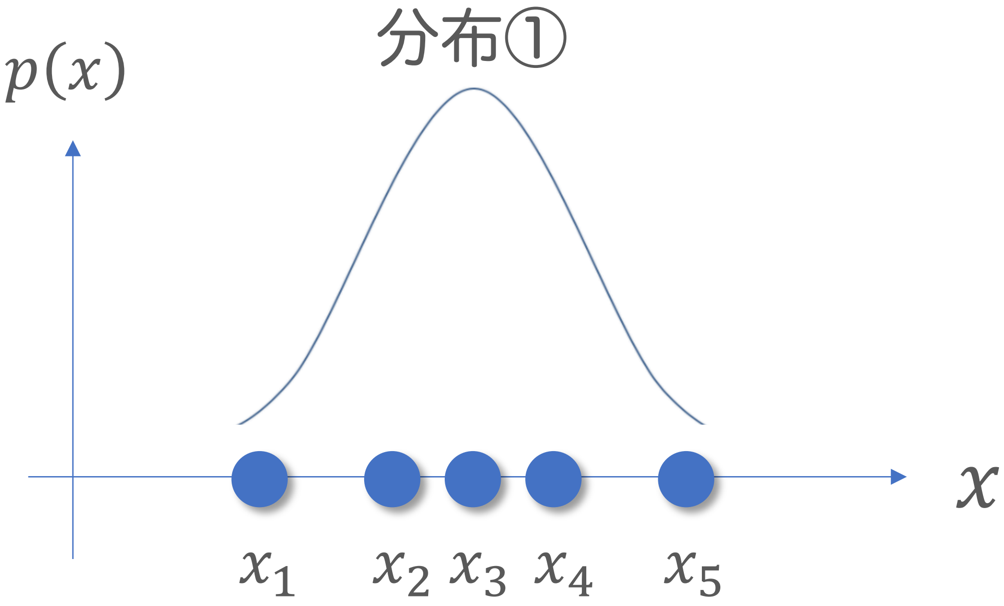
$$P(T = t_i|x_i) = y_i^{t_i}(1 - y_i)^{1-t_i} \quad y_i = \frac{1}{1 + e^{-wx_i}}$$

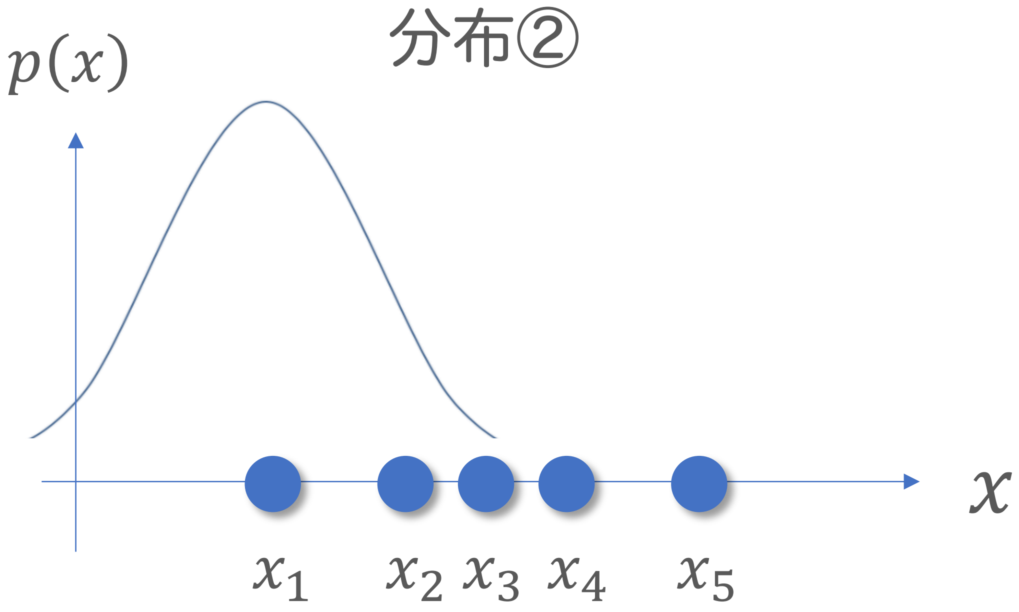
w の値をデータを元に推定する！

今回は最尤推定という手法を考えよう

最尤推定

所持しているデータを最も生成しそうな確率分布のパラメータを求める方法





（第３章理解確認問題（４）より）

ロジスティック回帰

以下のようなデータが生成される確率は？

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$\ldots$	x_N
t	1	1	1	0	1	1	1	1	$\ldots$	1

以下のようなデータが生成される確率は？

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	…	x_N
t	1	1	1	0	1	1	1	1	…	1

$$\prod_{n=1}^{N} y_i^{t_i}(1-y_i)^{1-t_i}$$

これを最大化する w が、

最尤推定という考え方における最も自然な w

この関数を尤度関数と呼ぶ

ロジスティック回帰

$$L(w) = \prod_{n=1}^{N} y_i^{t_i}(1 - y_i)^{1-t_i}$$

掛け算を多数回すると計算機が
アンダーフローを起こす可能性がある

ロジスティック回帰

$$L(w) = \prod_{n=1}^{N} y_i^{t_i}(1 - y_i)^{1-t_i}$$

掛け算を多数回すると計算機が
アンダーフローを起こす可能性がある

対数を取ることで和に変換
また最小化問題に置き換えるためにマイナスを付与

$$-\log L(w) = -\sum_{n=1}^{N}\{t_i \log y_i + (1 - t_i)\log(1 - y_i)\}$$

を最小化する w を最適化手法で求める

ロジスティック回帰

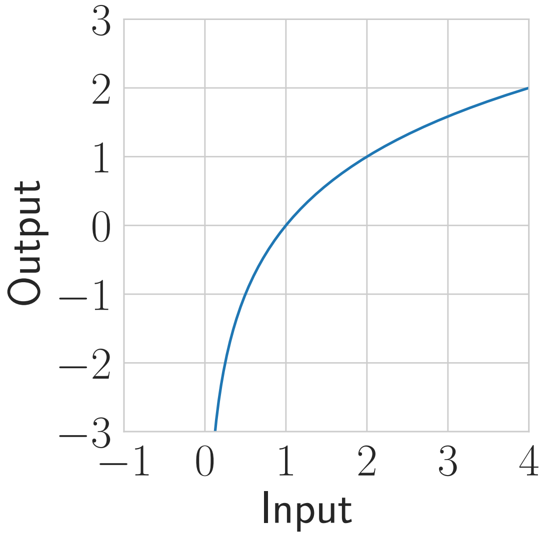
（１） $-\log \prod_{n=1}^{N} y_i^{t_i} (1-y_i)^{1-t_i}$ を計算し $-\sum_{n=1}^{N} \{t_i \log y_i + (1-t_i) \log(1-y_i)\}$ を導け

ヒント： $\log MN = \log M + \log N$ と $\log b^c = c \log b$ を利用する

（２）尤度関数に対して対数を取る操作を行なっているが

この操作は問題ないのだろうか？

ヒント：対数関数の概形



ロジスティック回帰

（3）以下のデータ及びロジスティック回帰モデル①, ②に対して尤度・負の対数尤度の値を計算せよ

ただし計算途中は小数第四位で四捨五入し、最終結果は小数第三位で四捨五入せよ

x 植物の身体の大きさ 標準化済み	t 種子を持つなら1 持たないなら0
1.0	1
0.9	1
-1.0	0
-0.9	0

① $y = \dfrac{1}{1+e^{-x}}$　　② $y = \dfrac{1}{1+e^{-10x}}$

計算には walframe alpha https://www.wolframalpha.com/ を利用せよ

例） $\frac{1}{1+e^{-10\times1.0}}$ なら `1 / (1 + e^(-10 * 1.0))` と入力して = をクリックし、「結果」を見る

例２） $\log 0.3$ なら `log(0.3)` と入力して = をクリックし、「結果」を見る

（1） $\log MN = \log M + \log N$ と $\log b^c = c \log b$ の性質を用いて計算すると

$$
\begin{aligned}
-\log \prod_{n=1}^{N} y_i^{t_i}(1-y_i)^{1-t_i} &= -\sum_{n=1}^{N} \log y_i^{t_i}(1-y_i)^{1-t_i} \\
&= -\sum_{n=1}^{N} \left\{\log y_i^{t_i} + \log(1-y_i)^{1-t_i}\right\} \\
&= -\sum_{n=1}^{N} \{t_i \log y_i + (1-t_i)\log(1-y_i)\}
\end{aligned}
$$

ロジスティック回帰

（2）

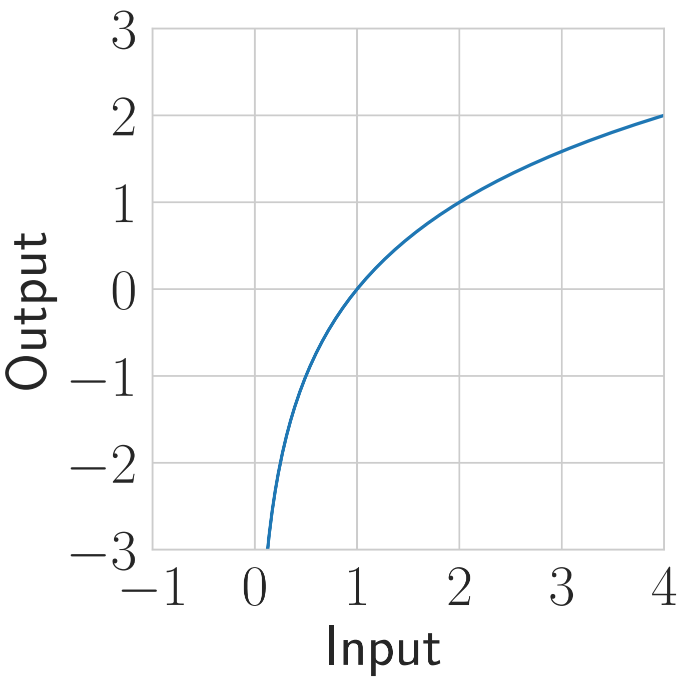
今回の場合は対数をとっても特に問題はない

今は尤度関数 $L(\boldsymbol{w})$ を最大化するパラメータ $\boldsymbol{w}$ を求めようとしている

対数関数は大小関係を保存する関数であるため

$L(\boldsymbol{w})$ を最大化する $\boldsymbol{w}$ は、$\log L(\boldsymbol{w})$ も最大化する

対数関数の概形



大小関係の保存

$$L(\boldsymbol{w}_1) > L(\boldsymbol{w}_2) \quad \Rightarrow \quad \log L(\boldsymbol{w}_1) > \log L(\boldsymbol{w}_2)$$

ロジスティック回帰

（３）

$$y = \frac{1}{1+e^{-x}}$$

x	t	y
1.0	1	0.731
0.9	1	0.711
−1.0	0	0.269
−0.9	0	0.289

$$L = 0.270 \ldots \approx 0.27$$
$$-\log L = 1.308 \ldots \approx 1.31$$

$$y = \frac{1}{1+e^{-10x}}$$

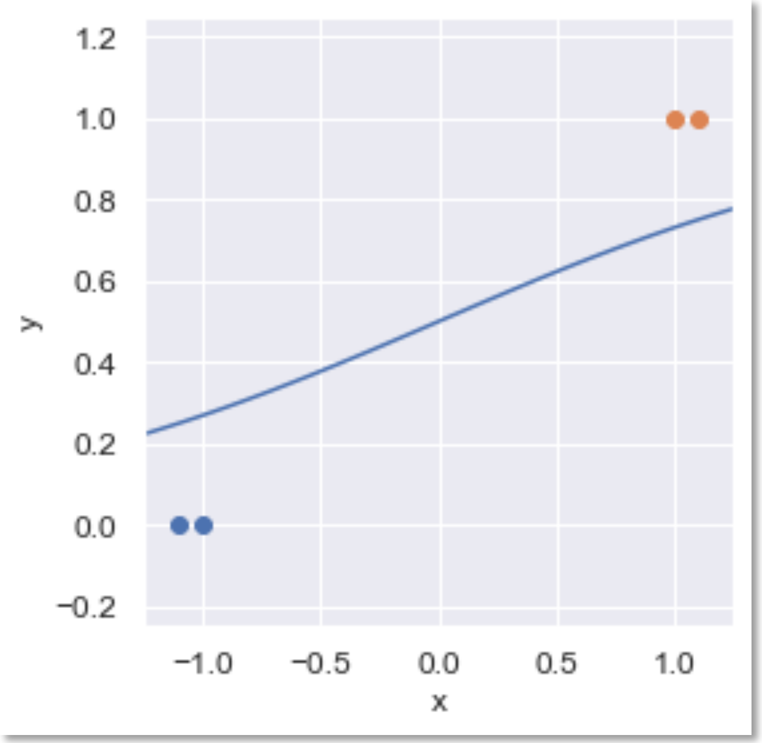
x	t	y
1.0	1	1.000
0.9	1	1.000
−1.0	0	0.000
−0.9	0	0.000

$$L = 1.000 \ldots \approx 1.00$$
$$-\log L = 0.000 \ldots \approx 0.00$$

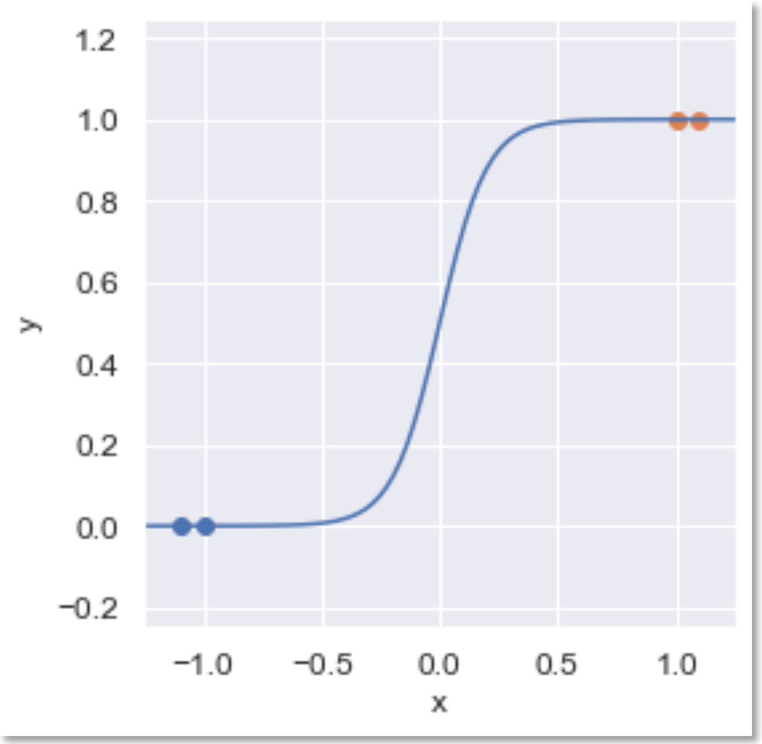
ロジスティック回帰

尤度の高いこちらの方が
良い w と言える

$$y = \frac{1}{1 + e^{-x}}$$



$$y = \frac{1}{1 + e^{-10x}}$$



ロジスティック回帰

補足）ロジスティック回帰は多次元の $\boldsymbol{x}$ にも容易に拡張可能

$$y = \frac{1}{1 + e^{-\boldsymbol{w} \cdot \boldsymbol{x}}} \qquad \boldsymbol{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix} \qquad \boldsymbol{x} = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{pmatrix}$$

終わりに

Day１ ：記述統計学の基礎

- 内容：統計量・可視化
- 修了演習：データの前処理技術（正規化と標準化）・箱ひげ図を用いた外れ値検出

Day２ ：確率

- 内容：確率の基礎・条件付き確率・ベイズの定理・独立
- 修了演習：ナイーブベイズによるスパムメール判定

Day３ ：確率分布

- 内容：離散型/連続型確率分布
- 修了演習：ロジスティック回帰

本講座はこれにて終了となります

Day1 ～ Day３ お疲れ様でした！

本講座で身に着けた確率・統計の知識は

機械学習/深層学習の習得に多いに役立つこと間違いなし！

またその他のデータ分析手法でも大活躍です

ぜひこれからも学習を続けてください！

付録

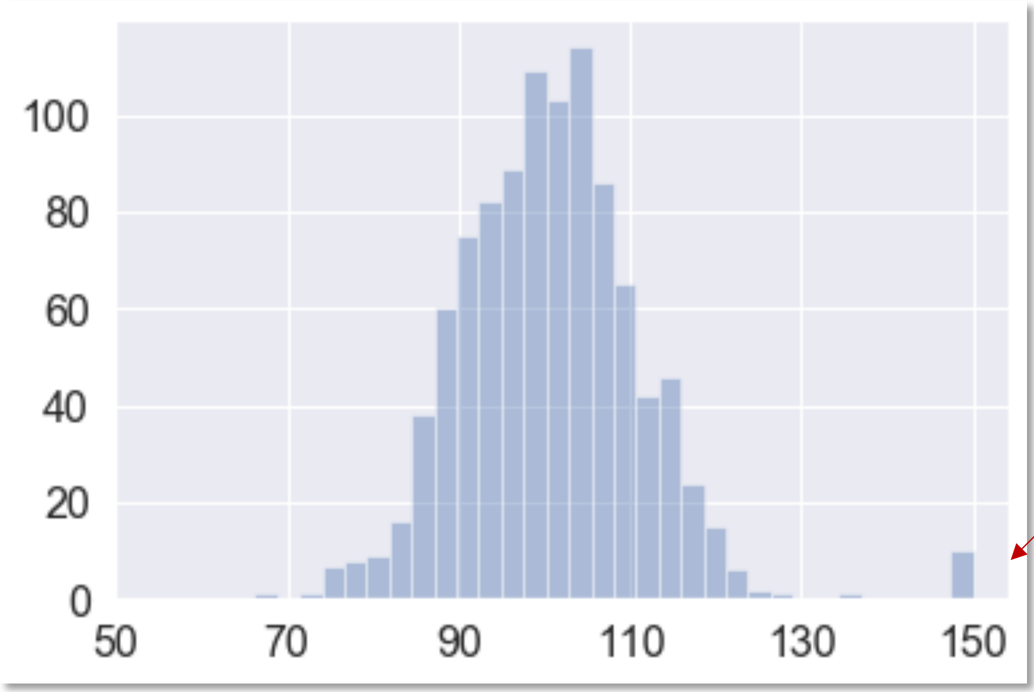
正規分布を用いた異常検知

異常値の可視化

データの異常値を確認するための可視化方法は？

→ 箱ひげ図・ヒストグラム

データ x のヒストグラム



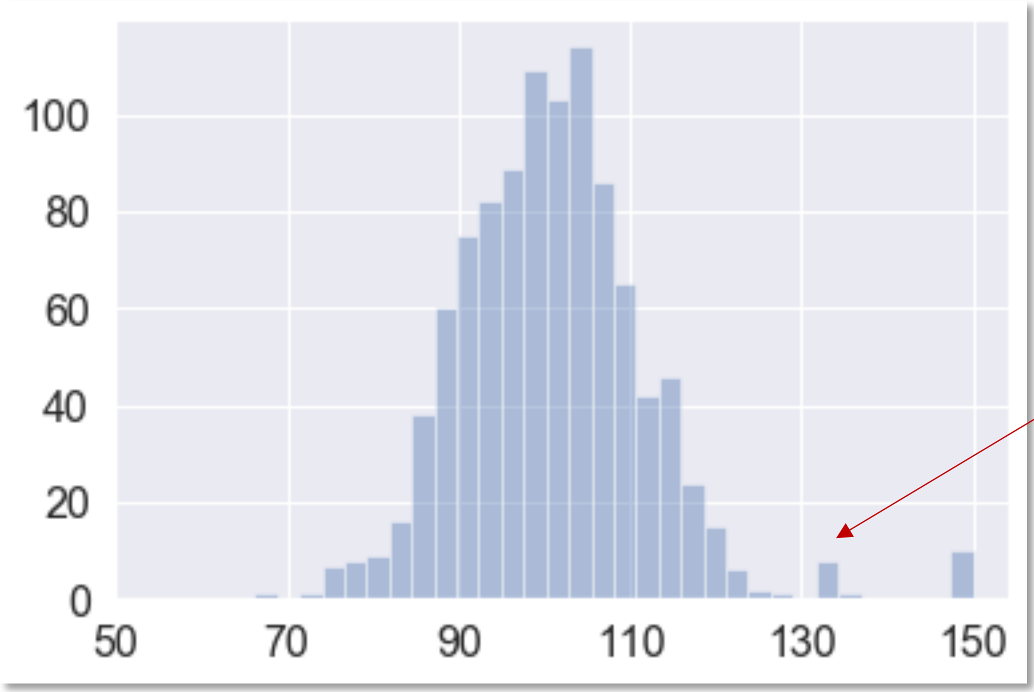
異常値に見える！

異常値の可視化

可視化に頼ると判断が難しいケースが出てくる

異常度合いを測るための定量的な指標が欲しい！ → ホテリング理論

データ x のヒストグラム

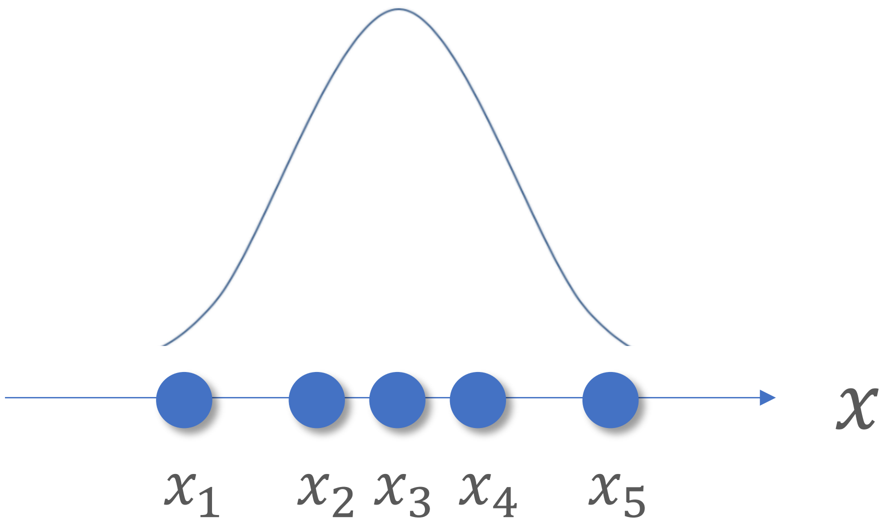


異常値...？

ホテリング理論による異常度の定量化

データ$x_1, x_2, \ldots, x_N$ は正規分布に従うと仮定

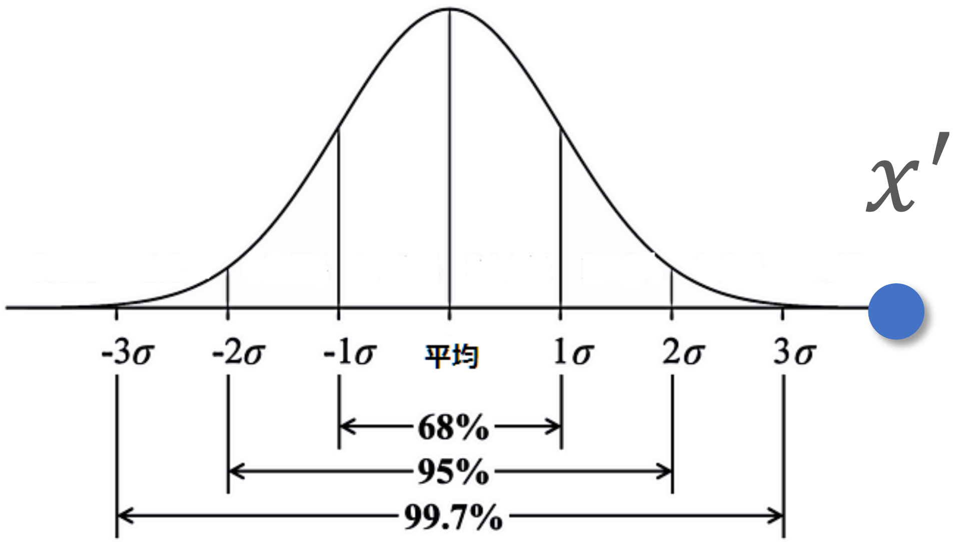
まずは μ と σ の値をデータを用いて推定する（最尤推定法など）



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

ホテリング理論による異常度の定量化

$p(x)$ を用いればデータの出やすさが分かる！



$p(x' \geq \mu + 3\sigma)$ 非常にレア！

異常値かも？

ホテリング理論では $p(x)$ にもう一工夫して異常度 $\alpha(x)$ を定義する

ホテリング理論による異常度の定量化

※ $\log x = \log_e x$

$p(x)$ を負の対数関数で変換

xに関係がない定数

$$-\log p(x) = -\log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \boxed{-\log \frac{1}{\sqrt{2\pi\sigma^2}}} + \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$\alpha(x) = \frac{(x-\mu)^2}{\sigma^2}$ これを x の異常度と定義

ホテリング理論の凄さは、この $\alpha(x)$ に対して理論的な拠り所を与えること

ホテリング理論による異常度の定量化

ホテリング理論

異常度 $\alpha(x)$ はデータ数 n が十分大きい時、自由度１のカイ二乗分布に従う

例）

自由度１カイ二乗分布において $\alpha(x) \geq 100$ となる確率は 0.016

これは非常に稀なので、その異常度を与える x は異常値と見なそう

と主張できる！

正規分布を用いた異常検知

（１）異常度 $\alpha(x)$ の導出を再度行え

（２） $\mu = 0, \sigma = 1$ のとき、 $x = 10$ の異常度 $\alpha(x = 10)$ を計算せよ

（３）ホテリング理論による異常検知の問題点について考えてみよ

正規分布を用いた異常検知

（１） $$-\log p(x) = -\log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = -\log \frac{1}{\sqrt{2\pi\sigma^2}} + \frac{1}{2} \boxed{\frac{(x-\mu)^2}{\sigma^2}}$$

$\alpha(x)$と定義

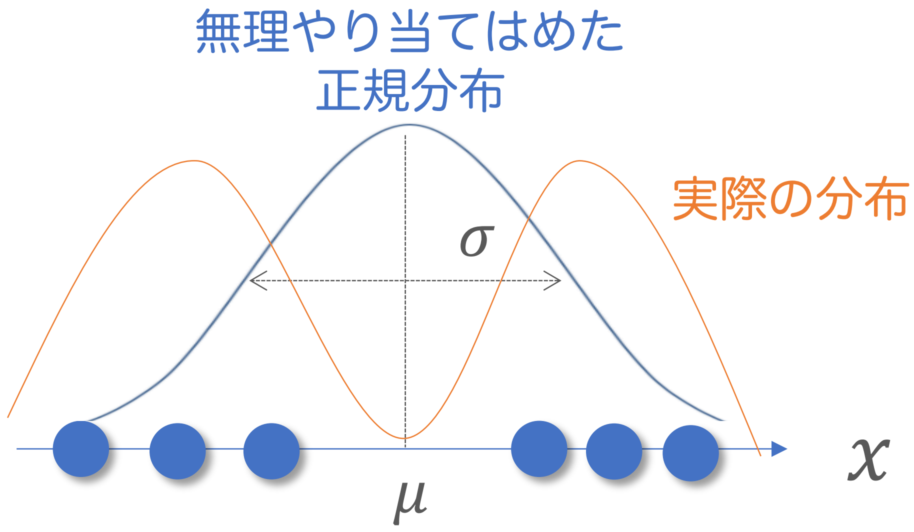
（２） $$\alpha(x' = 10) = \frac{(10-0)^2}{1^2} = 100$$

カイ二乗分布において 100 以上の値となる確率は 0.016

これは非常に小さいので異常値とみなしても良いだろう

正規分布を用いた異常検知

（３）　データが正規分布とは異なる分布に従う場合、異常値を正しく判定できない



正規分布のパラメータ μ, σ が時間変化していた場合には対応できない

$$\mu^{(t-1)}, \sigma^{(t-1)} \xrightarrow{\text{大きく変化する}} \mu^{(t)}, \sigma^{(t)}$$