

第8回マスフェスタ

確率の不思議と機械学習

鈴木 大慈

東京工業大学

情報理工学院数理・計算科学系

JSTさきがけ

2016/8/27

「確率」

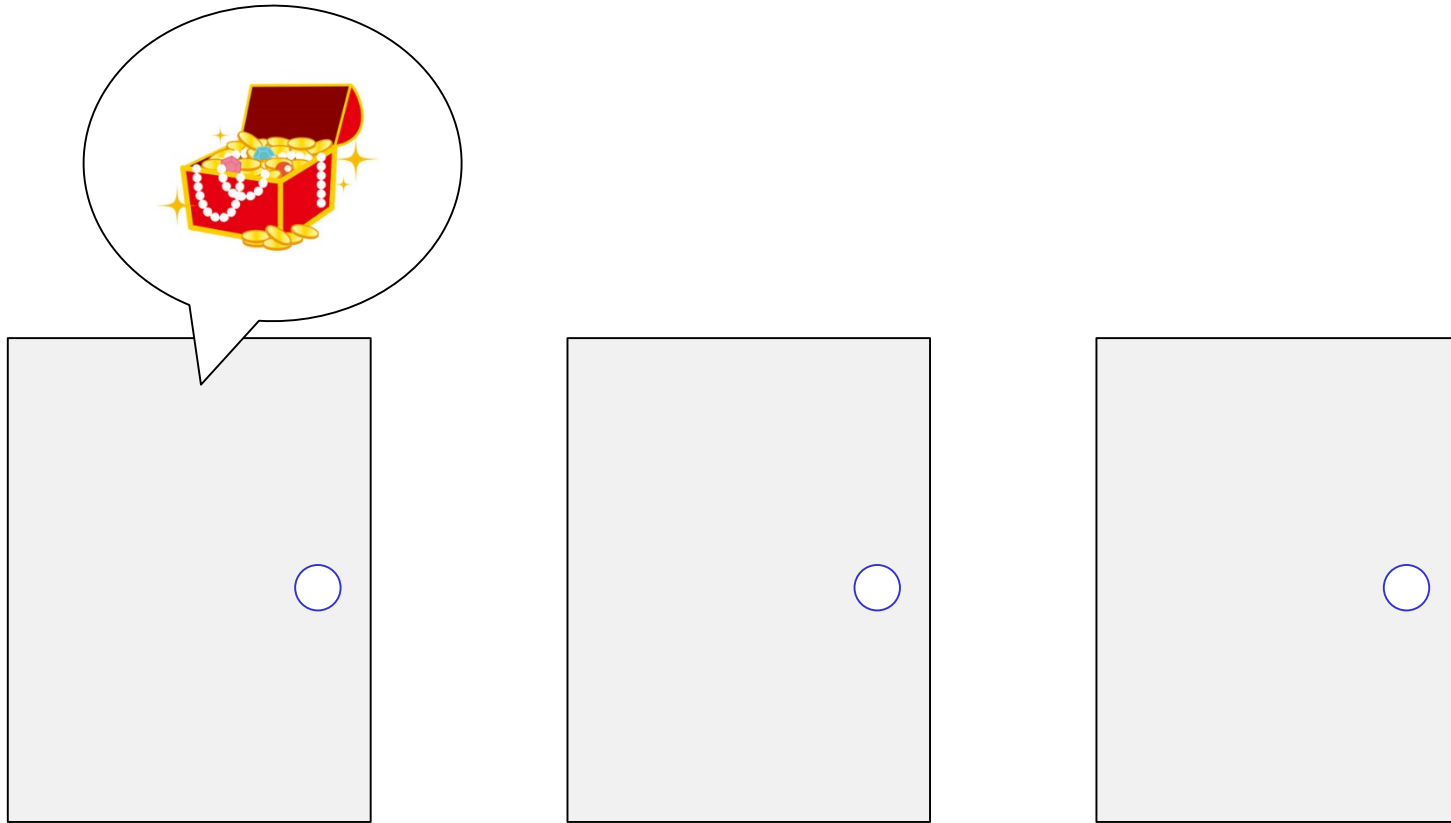


データを扱う方法
統計学・機械学習

確率にまつわる話

実験1

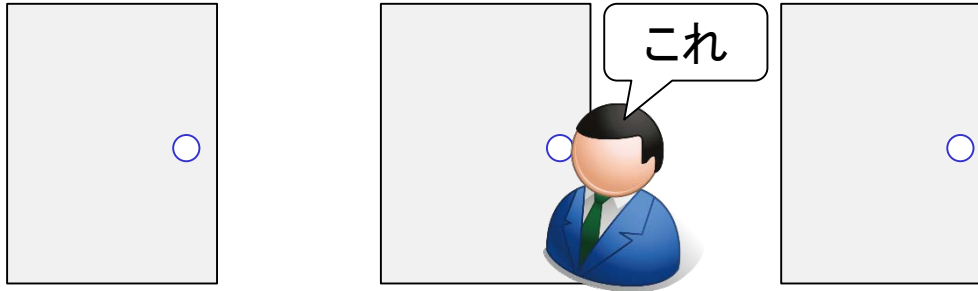
モンティホール問題



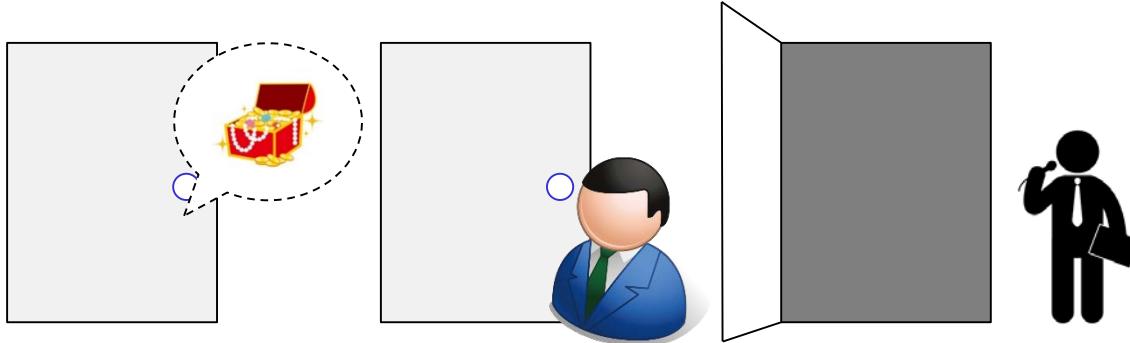
- 三つの扉のうちどこかに賞品が入っている.
- 挑戦者はそれを当てる.

ルール

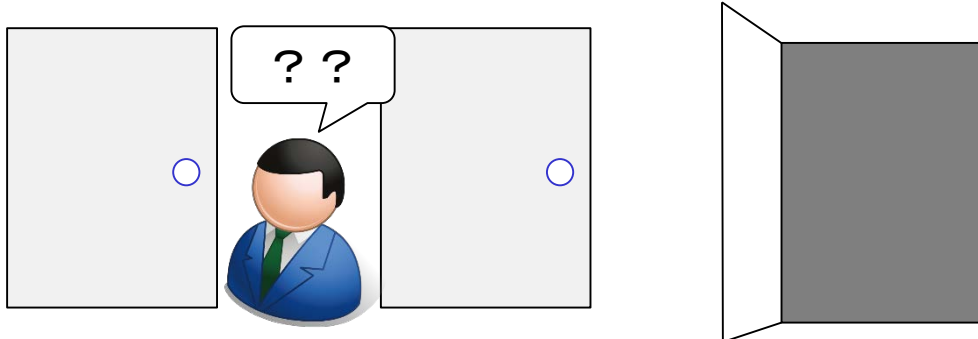
1. 挑戦者が一つの扉を指定する.



2. 司会者が残りの「正解でない扉」の一つを開ける。(どれを開けるかは等確立)



3. この結果を受けて挑戦者は選びなおす権利が与えられる.



扉を変えた方が得か？ 変えない方が得か？

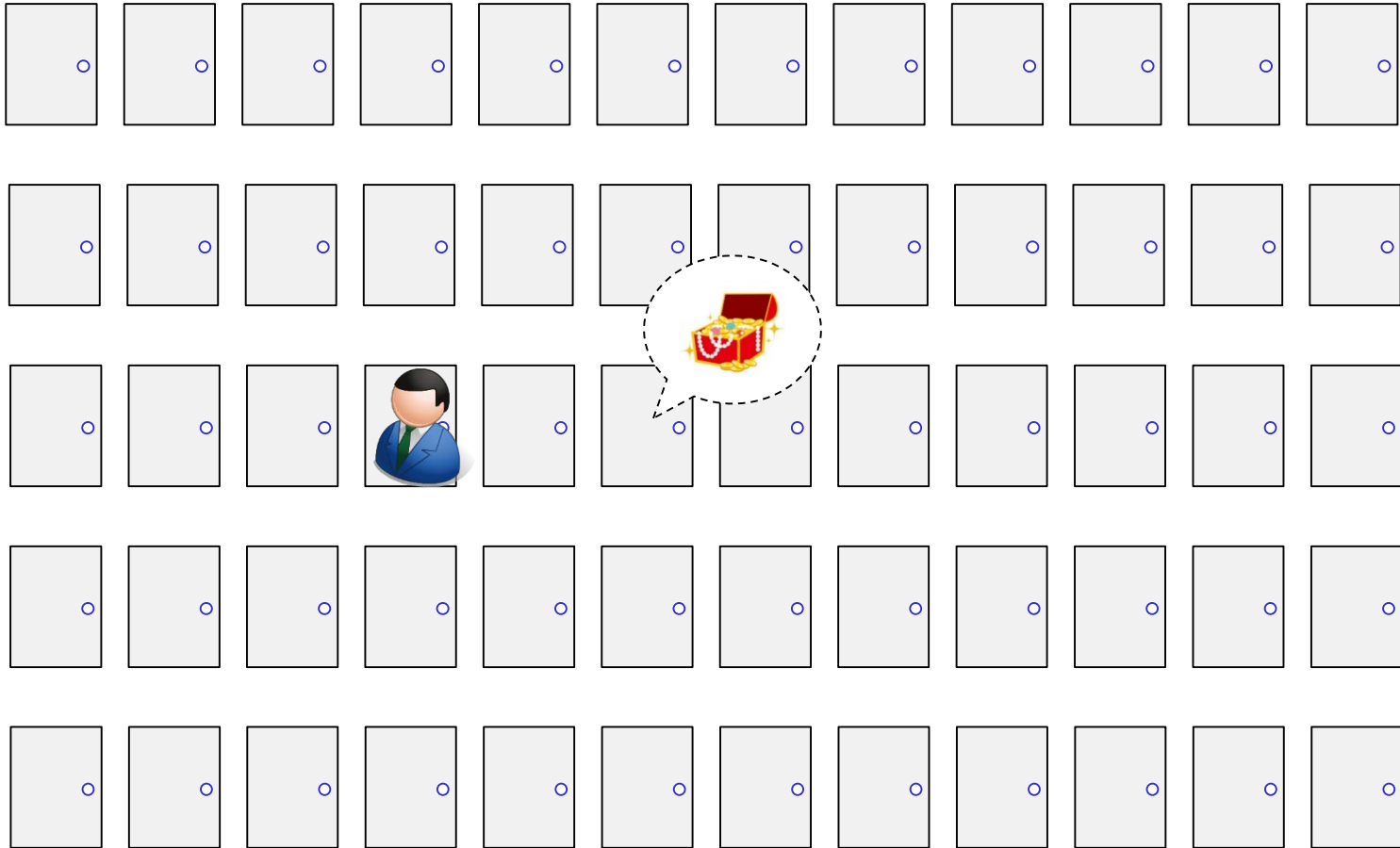
実際にやってみる

正解

選択を「変えた方」が得

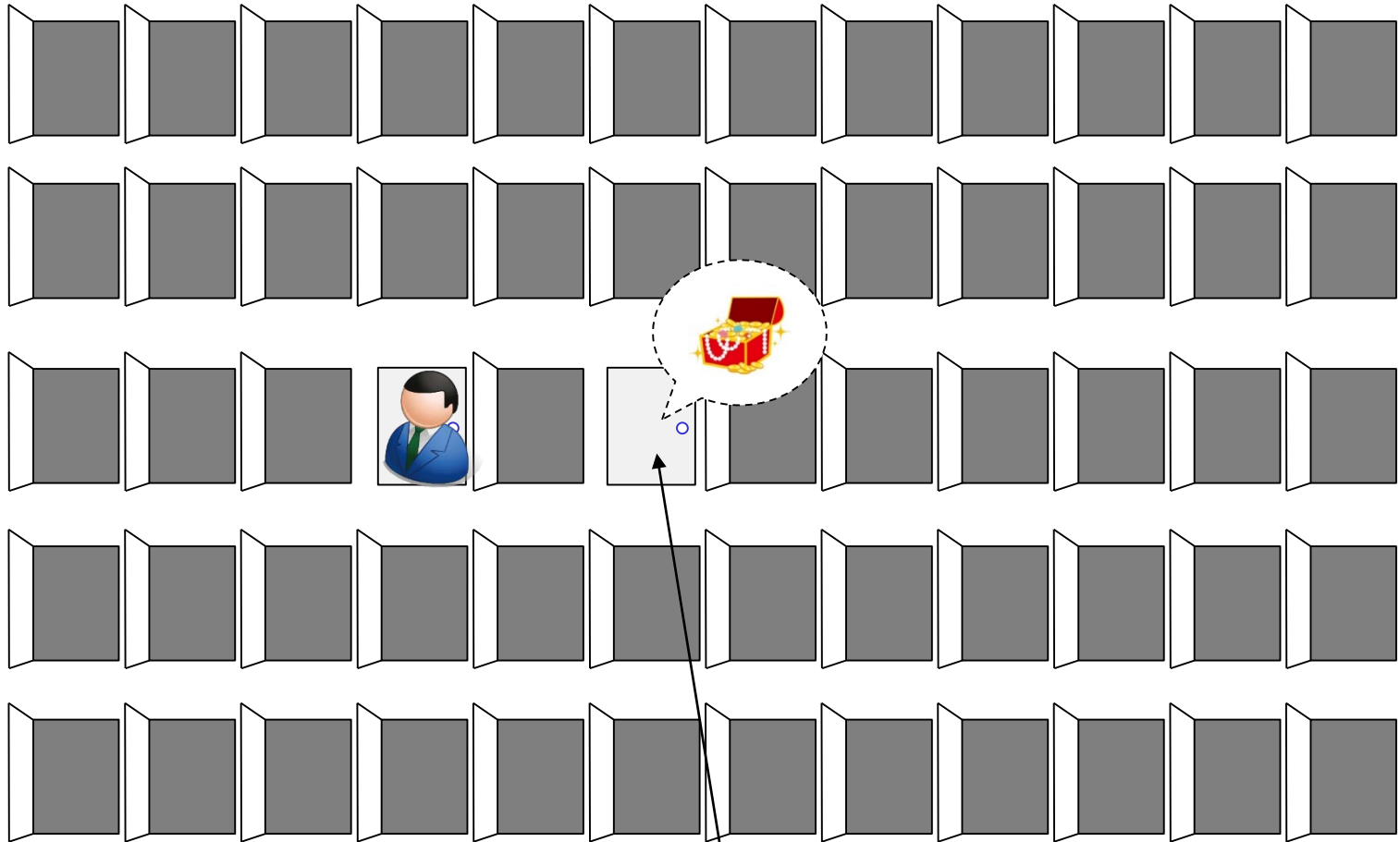
直感的な説明

扉の数を増やしてみよう.



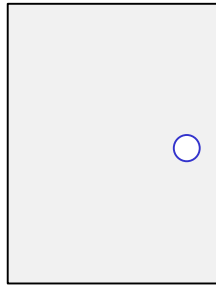
直感的な説明

司会者が残りの「正解でない扉」を開ける。

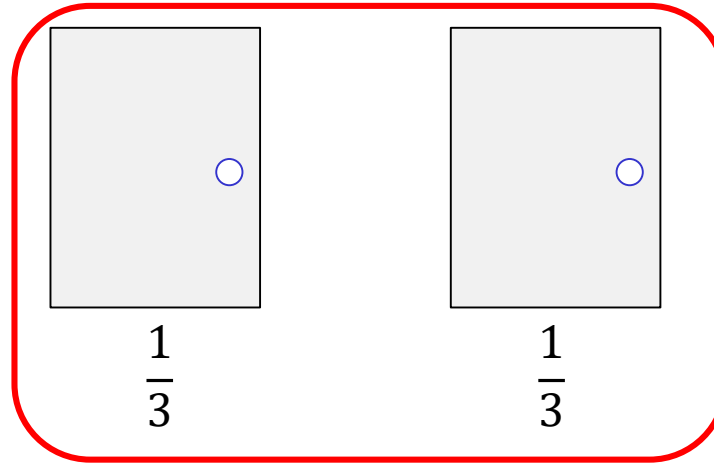


この扉がかなり怪しい

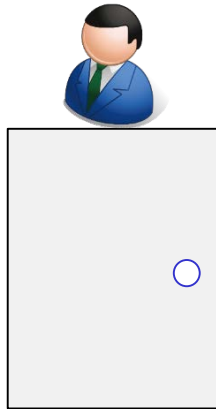
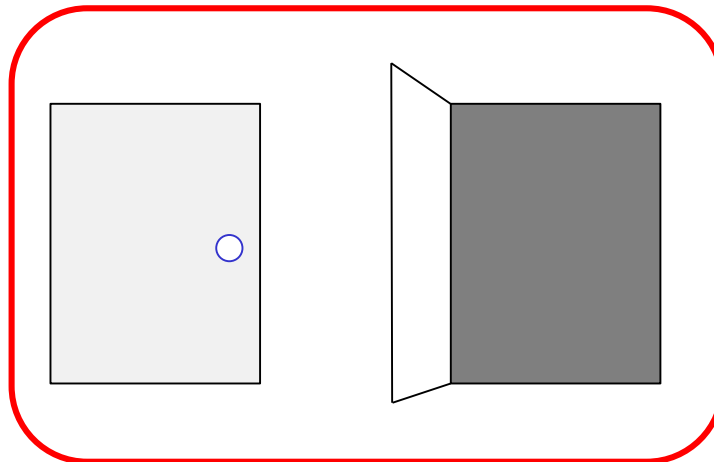
確率を計算


 $\frac{1}{3}$

正解の確率


 $\frac{1}{3}$
 $\frac{1}{3}$
 $\frac{2}{3}$

ドア2または3に
賞品がある確率
は2/3


 $\frac{1}{3}$

 $\frac{2}{3}$

ドア3の可能性はな
くなったので、ドア2
に賞品がある確率
は2/3

確率

そのままが正解:

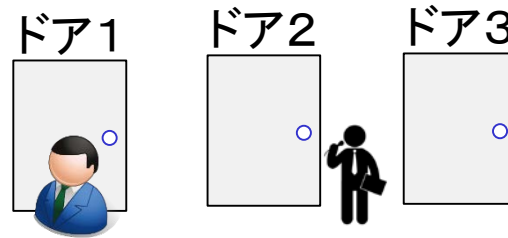
$\frac{1}{3}$

変えるが正解:

$\frac{2}{3}$

扉を開いても確率が残る

条件付き確率 (より厳密な導出)

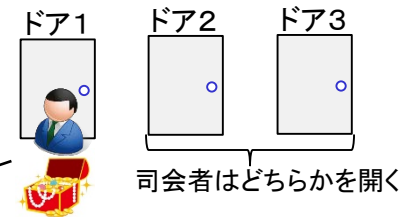


挑戦者はドア1を選び、司会者はドア2を開いたとしよう。

事象A: 司会者によってドア2が開かれる事象

事象B: ドア1に賞品がある事象

- $P(A)$: 事象Aが起きる確率
- $P(B)$: 事象Bが起きる確率
- $P(A|B)$: 事象Bで条件付けた事象Aの確率
「ドア1に賞品がある時に、ドア2が開かれる確率」
- $P(B|A)$: 「司会者によってドア2が開かれた時に、ドア1に賞品がある確率」
→ 求めたい確率 (扉を変えないことが正解である確率)



すぐわかること

$$P(B)=1/3,$$

$$P(A|B)=1/2 \text{ (司会者はドア2かドア3を開ける)}$$

$$P(A)=\text{少し難しい}$$

$$\begin{aligned}
 P(A) &= (\text{ドア1に賞品があつてかつドア2が開けられる確率}) + \\
 & \quad (\text{ドア2に賞品があつてかつドア2が開けられる確率}) + \\
 & \quad (\text{ドア3に賞品があつてかつドア2が開けられる確率}) \\
 &= P(\text{ドア2が開かれる} \mid \text{ドア1に賞品}) \times P(\text{ドア1に賞品}) + \\
 & \quad P(\text{ドア2が開かれる} \mid \text{ドア2に賞品}) \times P(\text{ドア2に賞品}) + \\
 & \quad P(\text{ドア3が開かれる} \mid \text{ドア3に賞品}) \times P(\text{ドア3に賞品})
 \end{aligned}$$

「ドア1に賞品があるという条件のもとでドア2が開かれる確率」という意味.

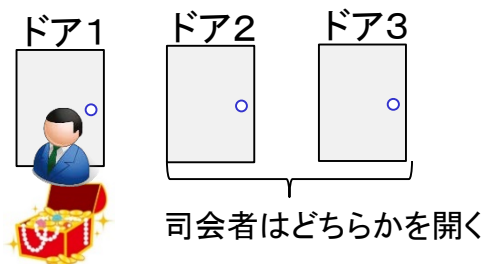
$$= \underbrace{\left(\frac{1}{3}\right)}_{\text{ドア1に賞品}} \times \underbrace{\left(\frac{1}{2}\right)}_{\text{ドア1に賞品があるならドア2と3のどちらかが等しい確率で開かれる}} + \quad \text{①}$$

$$\left(\frac{1}{3}\right) \times \underbrace{0}_{\text{賞品があるドアは開かれない}} + \quad \text{②}$$

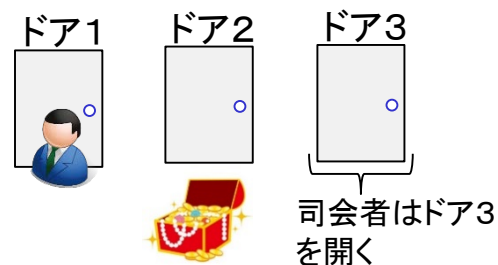
$$\underbrace{\left(\frac{1}{3}\right)}_{\text{ドア3に賞品}} \times \underbrace{(1)}_{\text{ドア3に賞品があればドア2が必ず開かれる}} \quad \text{③}$$

$$= \underline{\underline{1/2}}$$

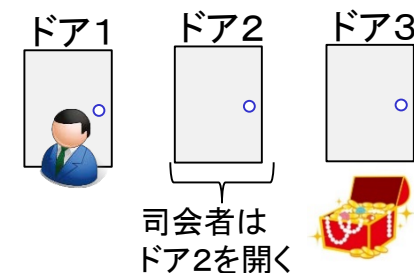
①



②



③

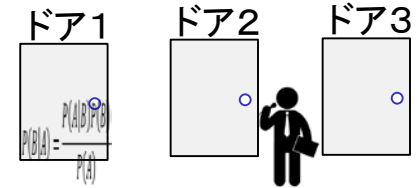


ベイズの定理

設定: 挑戦者はドア1を選び, 司会者はドア2を開いた。

事象A: 司会者によってドア2が開かれる事象

事象B: ドア1に賞品がある事象



わかったこと

$$P(B) = 1/3, P(A|B) = 1/2, P(A) = 1/2$$

ベイズの定理

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

「司会者がドア2を開いたという条件のもと, ドア1に賞品がある確率」

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

扉を変えない方が正解の確率

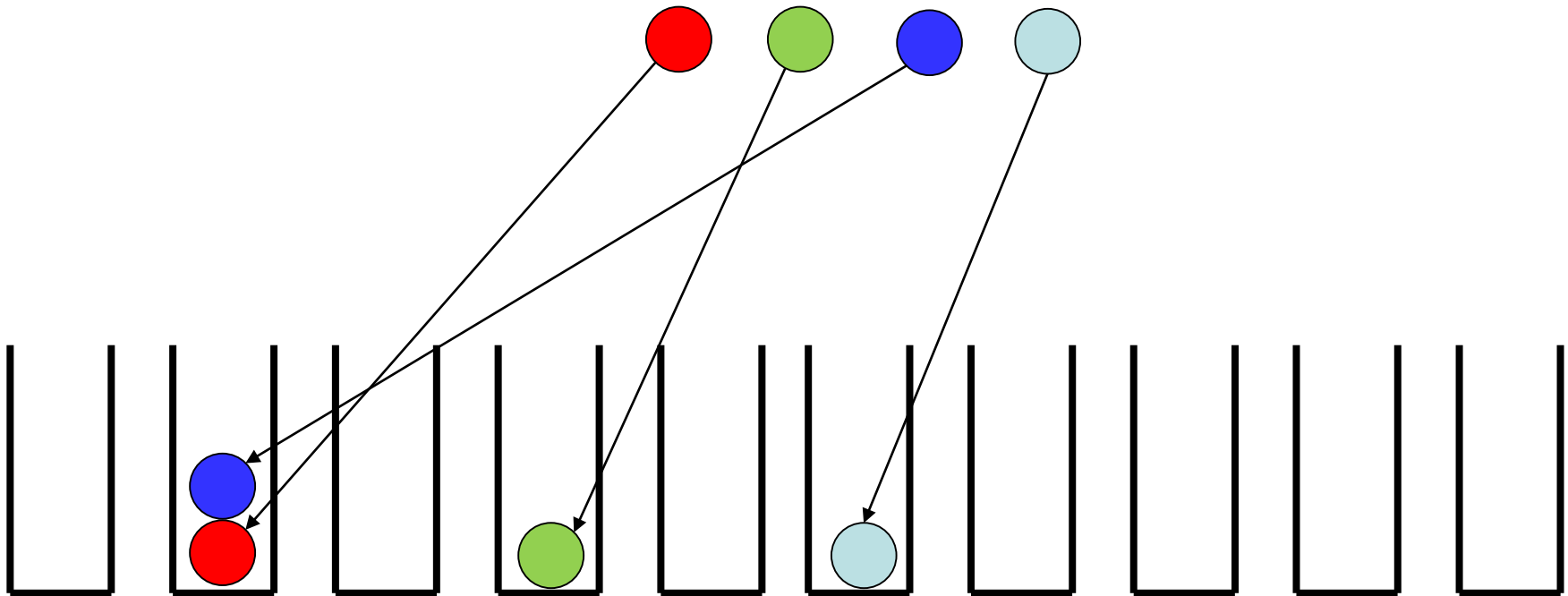
$$1 - P(B|A) = \frac{2}{3}$$

扉を変えた方が正解の確率

実験2

クーポンコレクター問題

10個のビンに1個ずつ球を入れてみよう。
どのビンに入れるかは等確率で選ぶ。



問題:

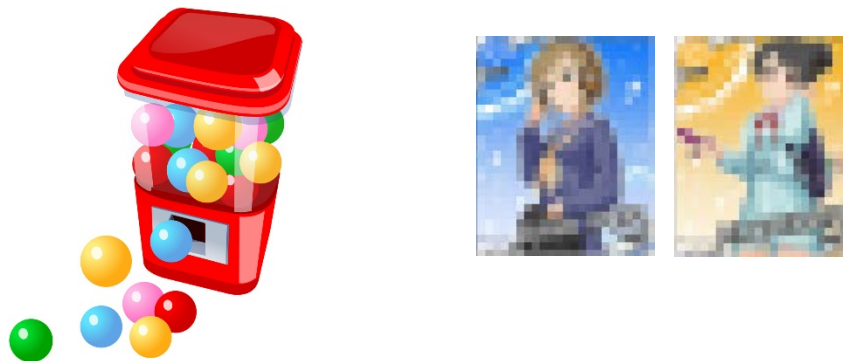
この試行を k 回(たとえば20回)繰り返した時に何個のビンに球が入っているか? 空のビンはあるか?

コンプガチャ問題

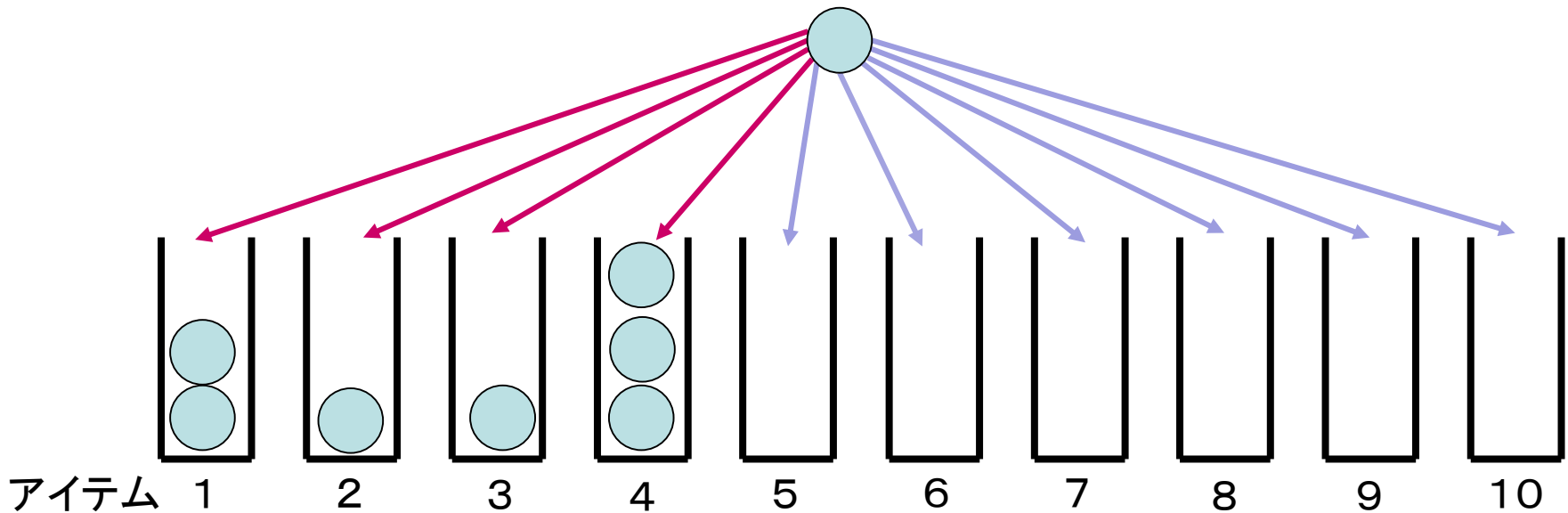
問題を少し修正：
何回球を落とせば、全部のビンに球が入るか？

言い換えてみる
「球を落とす→ガチャを引く」「ビン→アイテム」

何回ガチャを回せば、全アイテムをコンプリートできるか？



実際にためしてみよう



問題を分解:

4アイテムが手に入っているときに、次に5アイテム目が手に入るまでの回数
(k アイテムを手に入れてから $k + 1$ アイテム目を手に入れるまでの回数)

N_4 : 4アイテム→5アイテムまでにガチャを回した回数

$$N_4 = x \text{ である確率} \quad P(N_4 = x) = \left(\frac{4}{10}\right)^{x-1} \cdot \frac{6}{10}$$

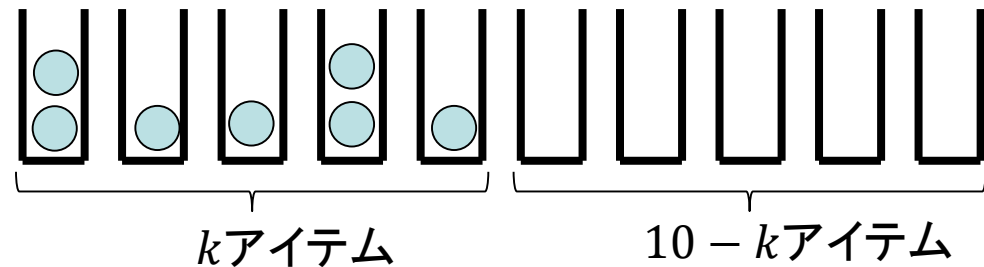
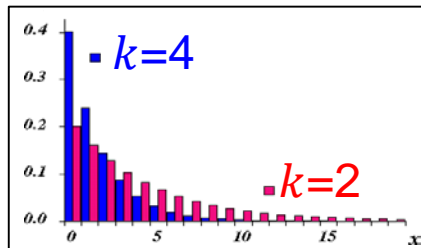
最初の $x - 1$ 回はすでに手に入っている4アイテムのうちどれかが出て、最後の x 回目に残りの6アイテムのうちどれかが出る確率

一般化

N_k : k アイテムを手に入れてから $k + 1$ アイテム目を手に入れるまでの回数

$$N_k = x \text{ である確率} \quad P(N_k = x) = \left(\frac{k}{10}\right)^{x-1} \cdot \frac{10-k}{10}$$

最初の $x - 1$ 回はすでに手に入っている k アイテムのうちどれかが出て、最後の x 回目に残りの $10 - k$ アイテムのうちどれかが出る確率



N_k の平均(期待値)

$$\sum_{x=1}^{\infty} x P(N_k = x) = \sum_{x=1}^{\infty} x \left(\frac{k}{10}\right)^{x-1} \cdot \frac{10-k}{10} = ?$$

$$\begin{aligned}
\sum_{x=1}^{\infty} xP(N_k = x) &= \sum_{x=1}^{\infty} x \left(\frac{k}{10}\right)^{x-1} \cdot \frac{10-k}{10} \\
&= \sum_{x=1}^{\infty} \frac{d \left[\left(\frac{k}{10}\right)^x \right]}{dk} \cdot (10-k) && \text{Kについての微分に置き換え} \\
&= \left[\frac{d}{dk} \sum_{x=1}^{\infty} \left(\frac{k}{10}\right)^x \right] \cdot (10-k) \\
&= \frac{d}{dk} \left(\frac{k/10}{1-k/10} \right) \cdot (10-k) && \text{等比級数の和の公式} \\
&= \left(\frac{1/10}{1-k/10} + \frac{k/100}{(1-k/10)^2} \right) \cdot (10-k) = \frac{10}{10-k}
\end{aligned}$$

9アイテム目をゲットした時から10アイテム目をゲットするまでの平均回数は10回！

コンプリートするまでの回数

コンプリートするまでの回数

$$N_0 + N_1 + N_2 + \cdots + N_9 = \sum_{k=0}^9 N_k$$

コンプリートするまでの回数の平均 = N_k の平均の和

$$\sum_{k=0}^9 \frac{10}{10-k} = 10 \underbrace{\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{10} \right)}_{\text{調和数(約2.93)}} \cong 29.3$$

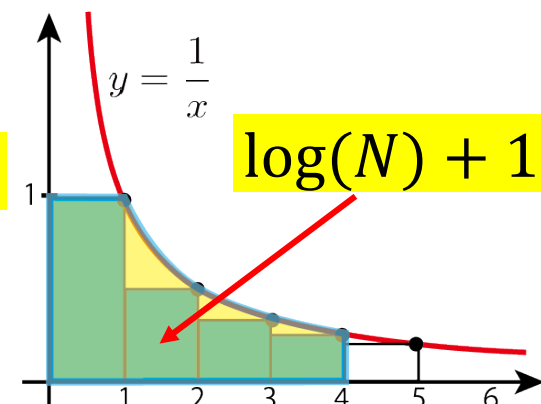
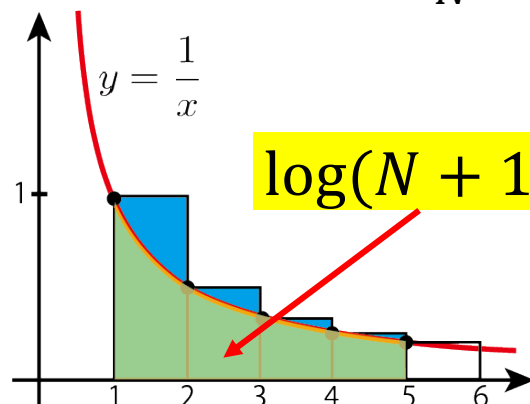
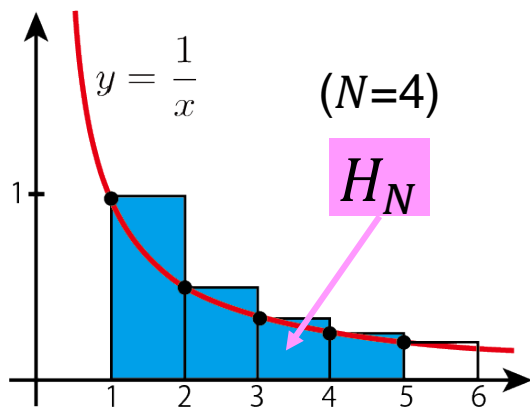
【一般化】 N 種類のアイテムをコンプリートするまでの平均回数

$$\sum_{k=0}^{N-1} \frac{N}{N-k} = N \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N} \right)$$

調和数の大きさを評価

【一般化】 N 種類のアイテムをコンプリートするまでの平均回数

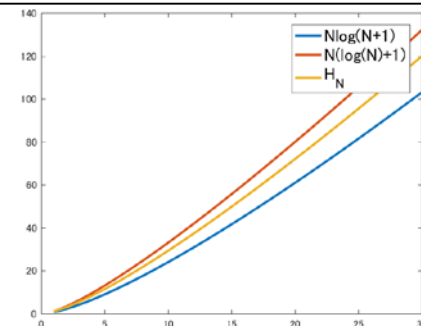
$$\sum_{k=0}^{N-1} \frac{N}{N-k} = N \underbrace{\left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{N} \right)}_{H_N}$$



$$\log(N+1) \leq H_N \leq \log(N) + 1$$

$N \log(N+1) \leq \text{平均回数} \leq N(\log(N) + 1)$

N	平均回数
30	119.8
100	518.7



実験3

コイン投げ

-確率から統計へ-

実験

- コイン投げ

10回コインを投げて表が出た回数を数えてください。

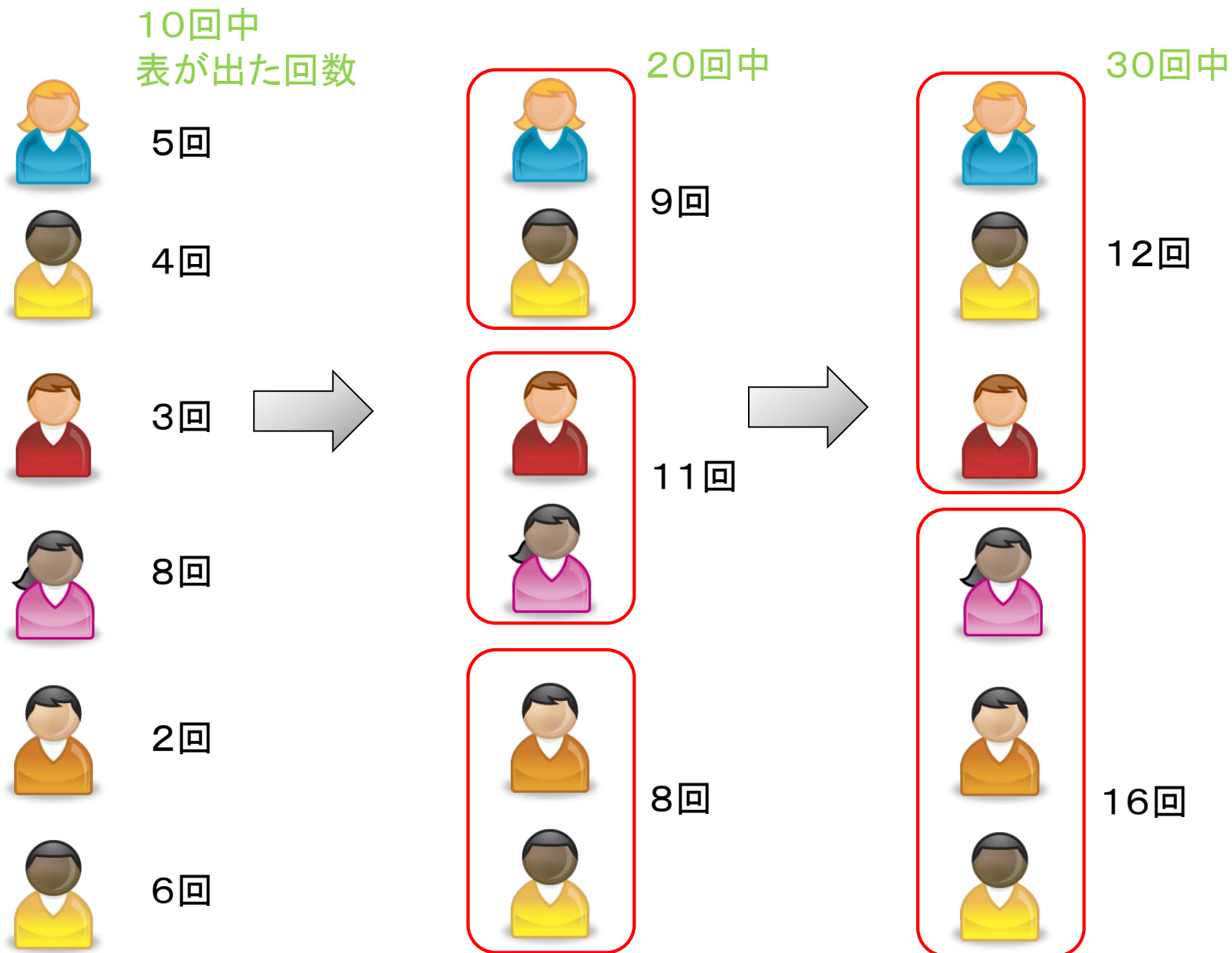


表が出た回数

0	1	2	3	4	5	6	7	8	9	10

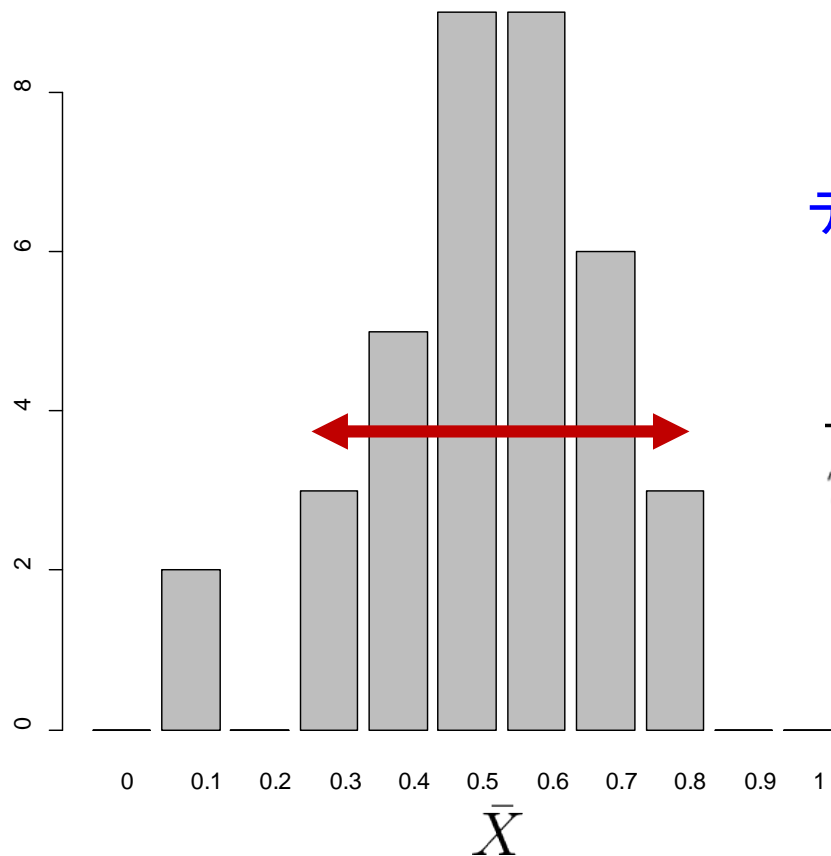
人数

グループ分けして平均



分散

Rでプロット



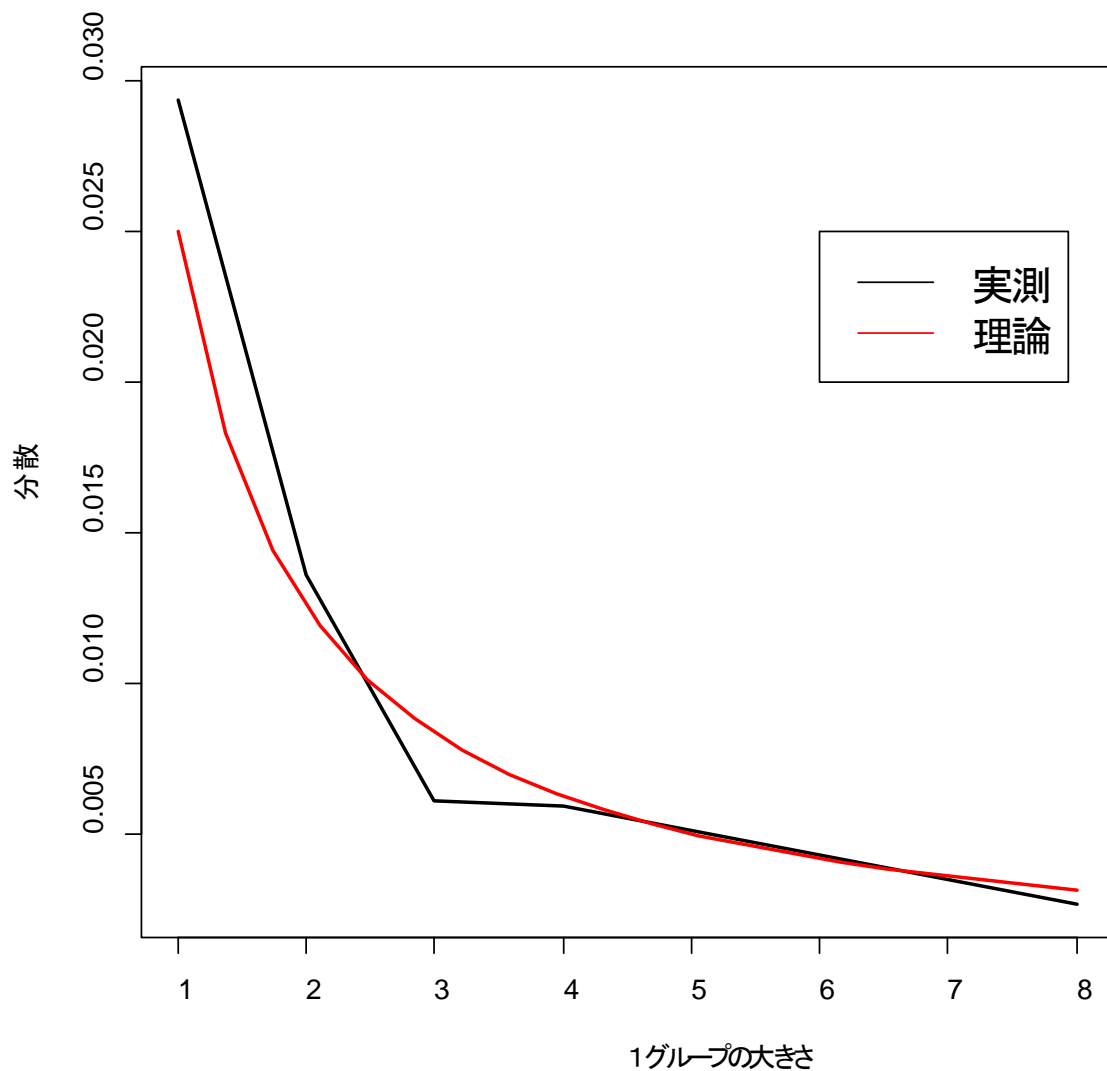
データのバラツキ

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

X_i : グループ内で表が出た割合

\bar{X} : 全体平均

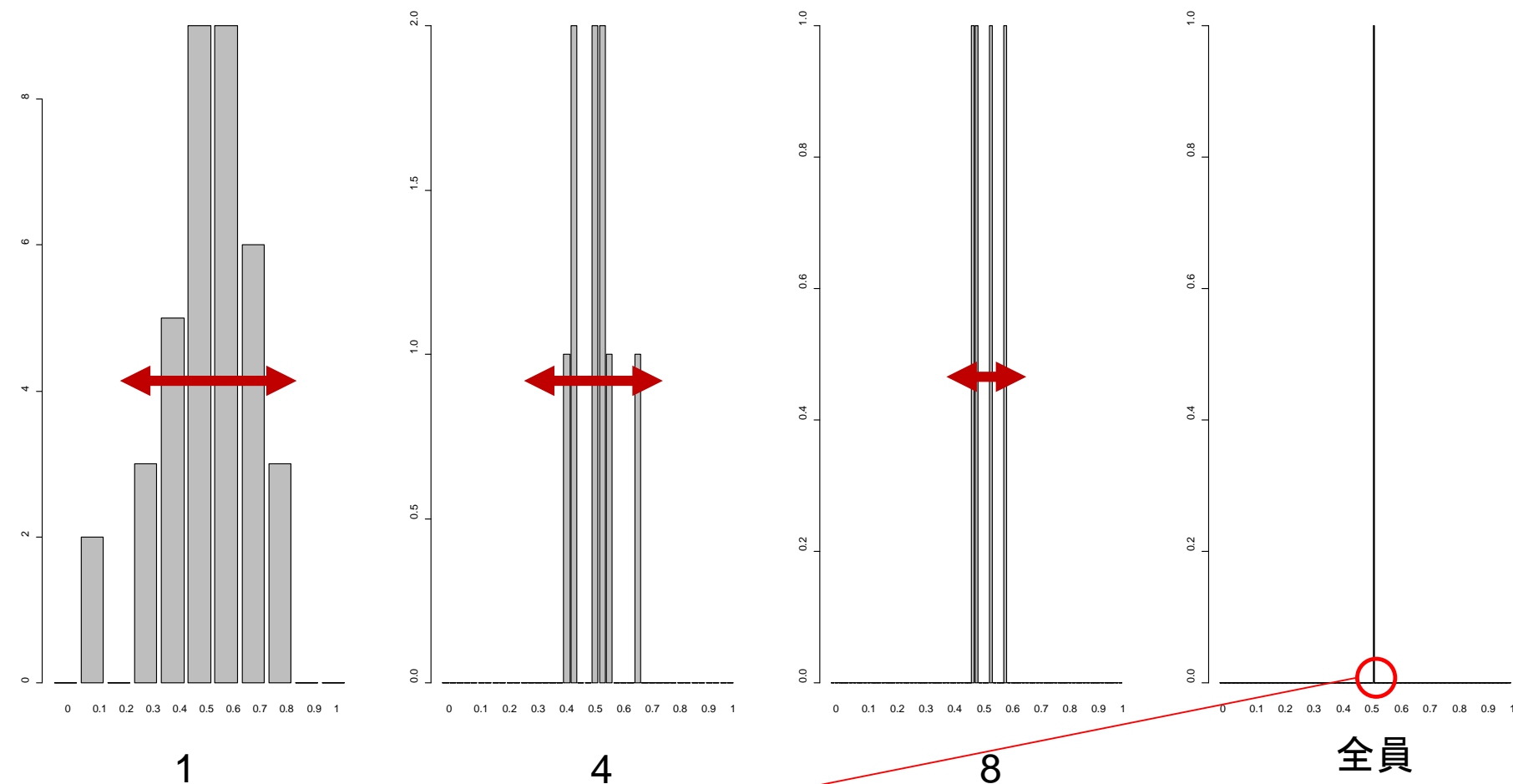
分散の理論値



グループのサイズを大きくすると分散が小さくなってゆく。

理論値 $\frac{1}{4N}$

ヒストグラムの変遷



収束した値は？ → 0.5に近いはず

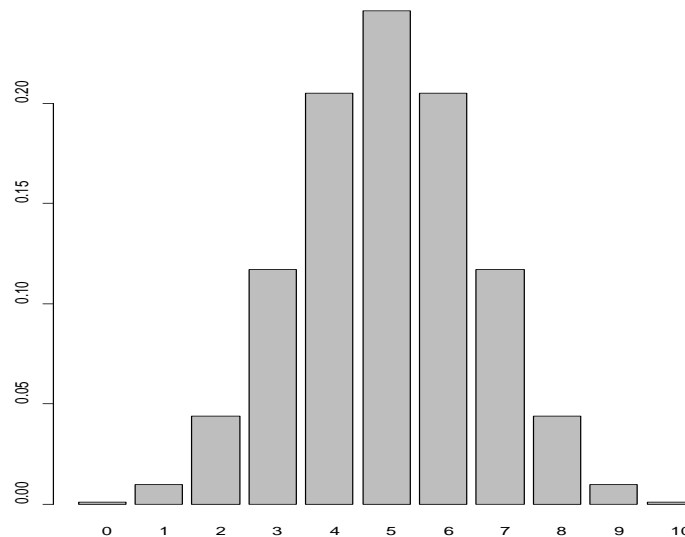
なぜ？

コイン投げ＝「二項分布」

n回コインを投げてx回表が出る確率

$$P(x) = \frac{n!}{x!(n-x)!} \left(\frac{1}{2}\right)^n$$

n=10

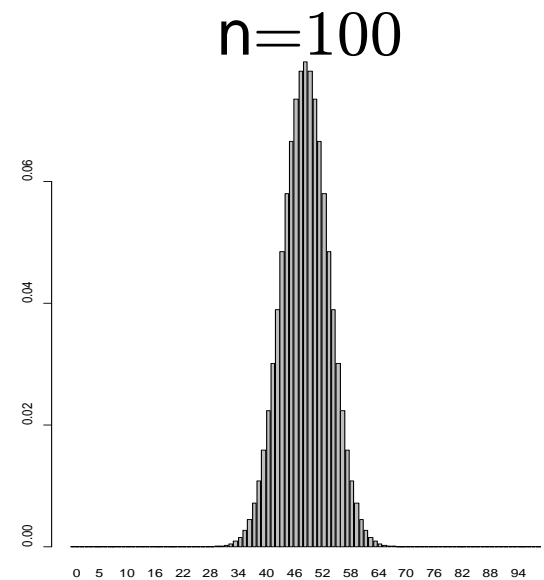
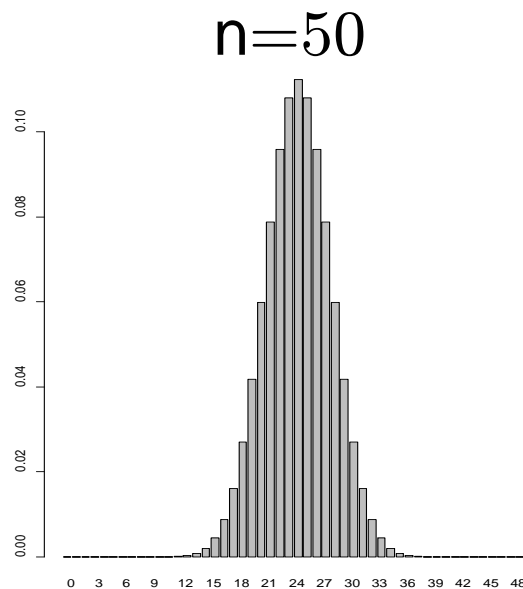
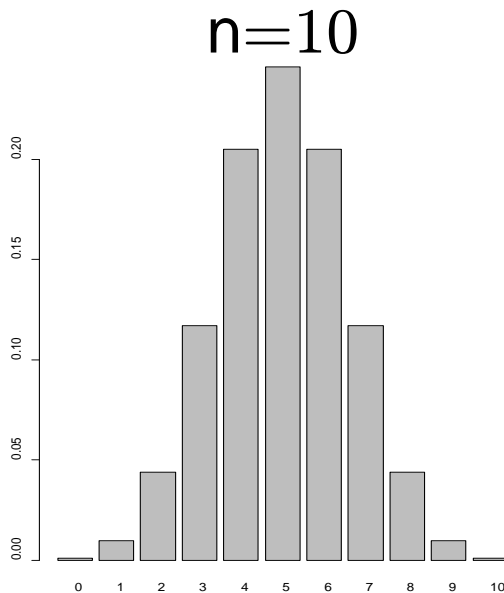


なぜ？

コイン投げ＝「二項分布」

n回コインを投げてx回表が出る確率

$$P(x) = \frac{n!}{x!(n-x)!} \left(\frac{1}{2}\right)^n$$



大数の法則

n回コインを投げて
表だった回数の割合

$$\xrightarrow{n \rightarrow \infty} \frac{1}{2}$$

実際，教室全体でコインを投げた平均値はほぼ0.5
(のはず)

では，有限のnでは0.5からどれだけ離れうる？

中心極限定理

表が出た割合は**正規分布**で近似できる

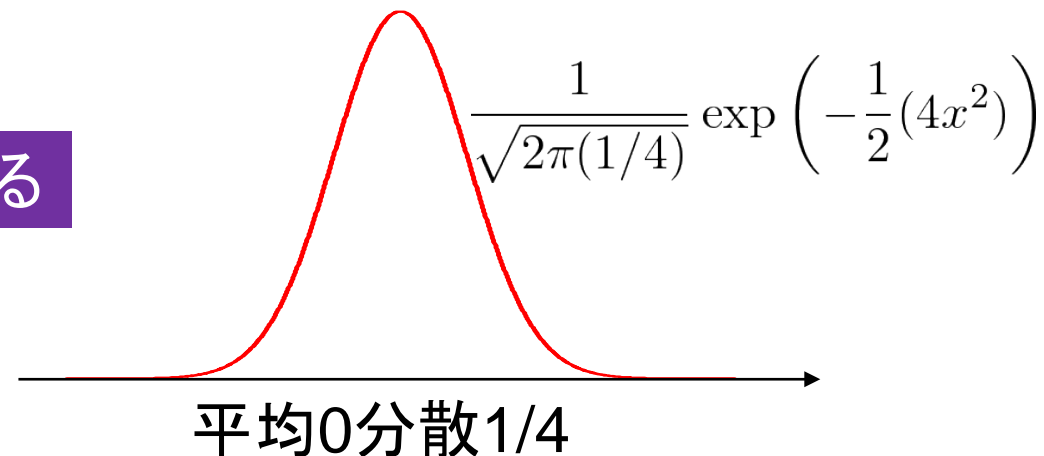
n回コインを投げて表が出た割合： Z_n

※ 大数の法則より Z_n は0.5に収束

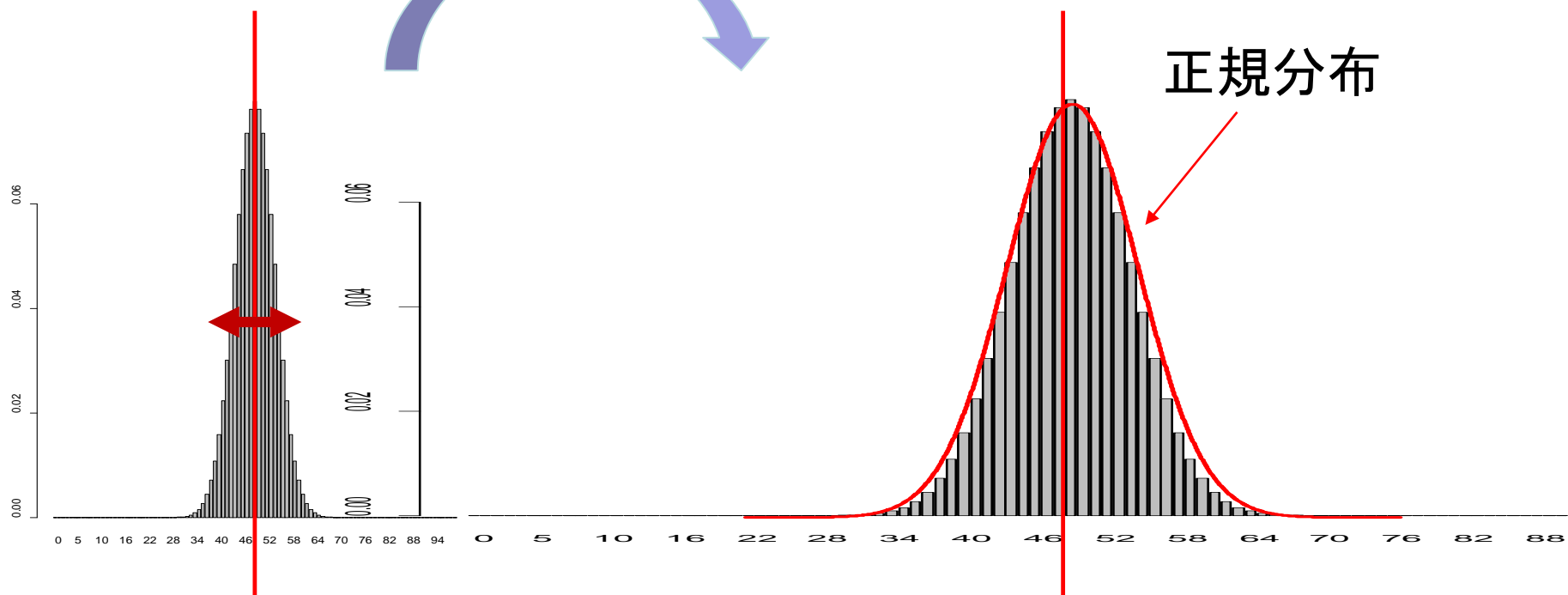
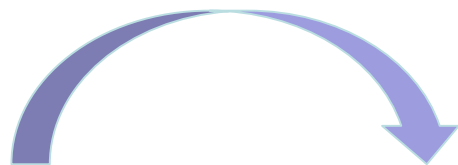
収束先との差を \sqrt{n} 倍して膨らみます。

$\sqrt{n}(Z_n - 0.5) \longrightarrow$ 正規分布

平均からの誤差が見積もれる



\sqrt{n} 倍して膨らみます.

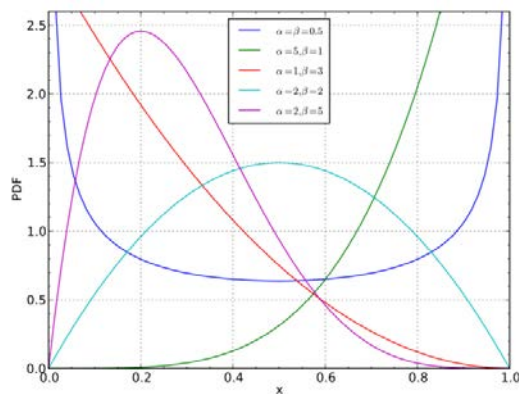


$n=100$

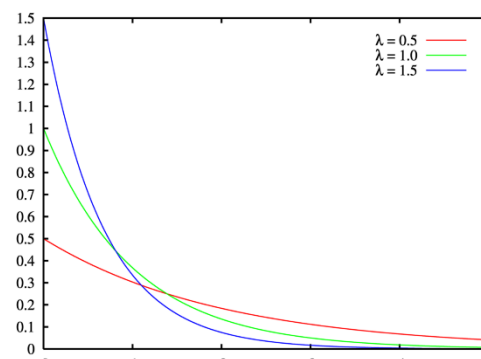
中心極限定理【一般形】

実はあらゆる確率変数においてサンプル平均は正規分布で近似できる。

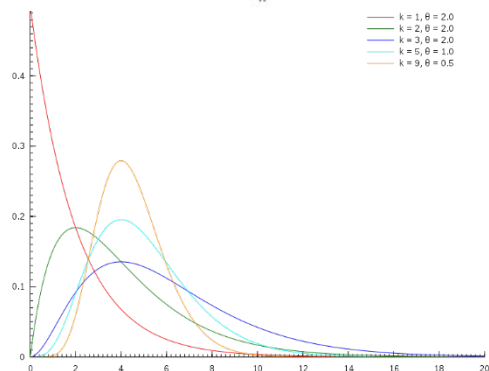
世の中にはいろいろな分布がある



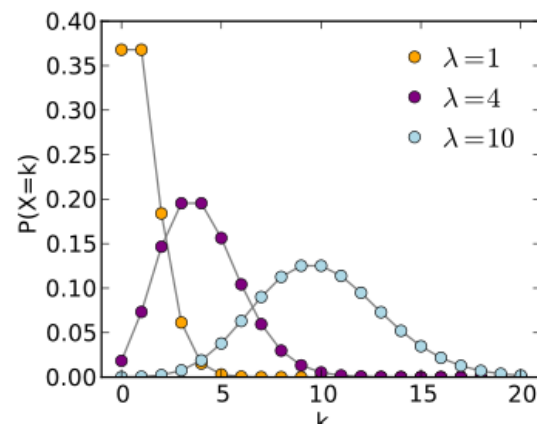
ベータ分布



指数分布



ガンマ分布



ポアソン分布

中心極限定理【一般形】

実はあらゆる確率変数においてサンプル平均は正規分布で近似できる。

$$X_1, \dots, X_n$$

ある分布からたくさんサンプルを得る。
(コイン投げだと、たくさんコインを投げる)

$$\sqrt{n} \left(\underbrace{\frac{1}{n} \sum_{i=1}^n X_i}_{\text{サンプル平均}} - \underbrace{\mu}_{\text{本当の平均}} \right)$$

サンプル平均 本当の平均

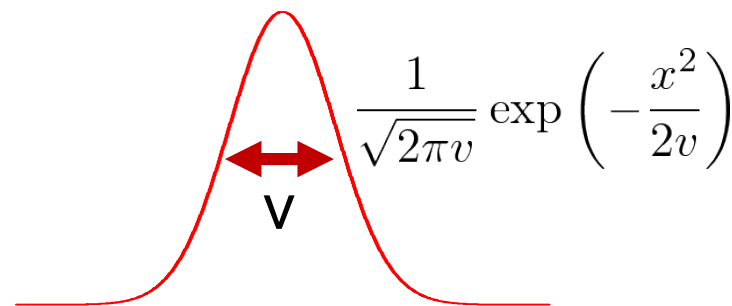
普遍性

→

正規分布

(平均0分散 v)

v は1サンプルの分散

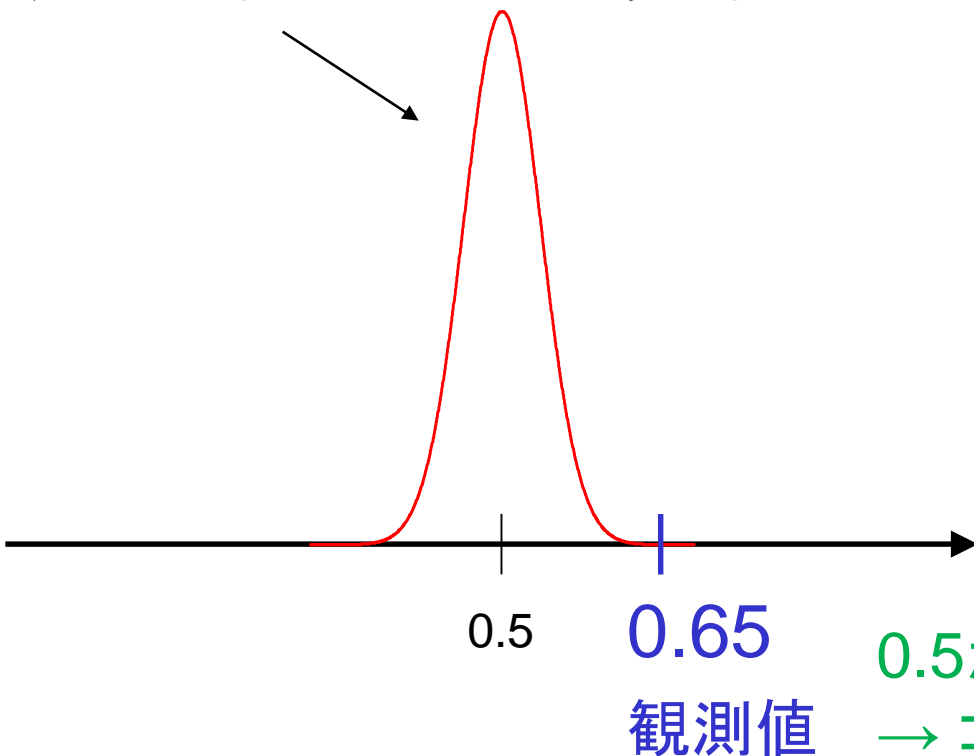


Rで確認

中心極限定理の使い方

- コイン投げのイカサマを見抜く

表がでる割合の確率(ほぼ正規分布)



平均値から0.15以上離れる確率

n	確率
10	0.34
100	0.0027
1000	$2 \cdot 10^{-21}$
10000	10^{-198}

定量的に「ありえなさ度合」
がはかれる！

0.5から離れている
→ コインが歪んでる？
それとも偶然？

仮説検定

「当選確実」の謎

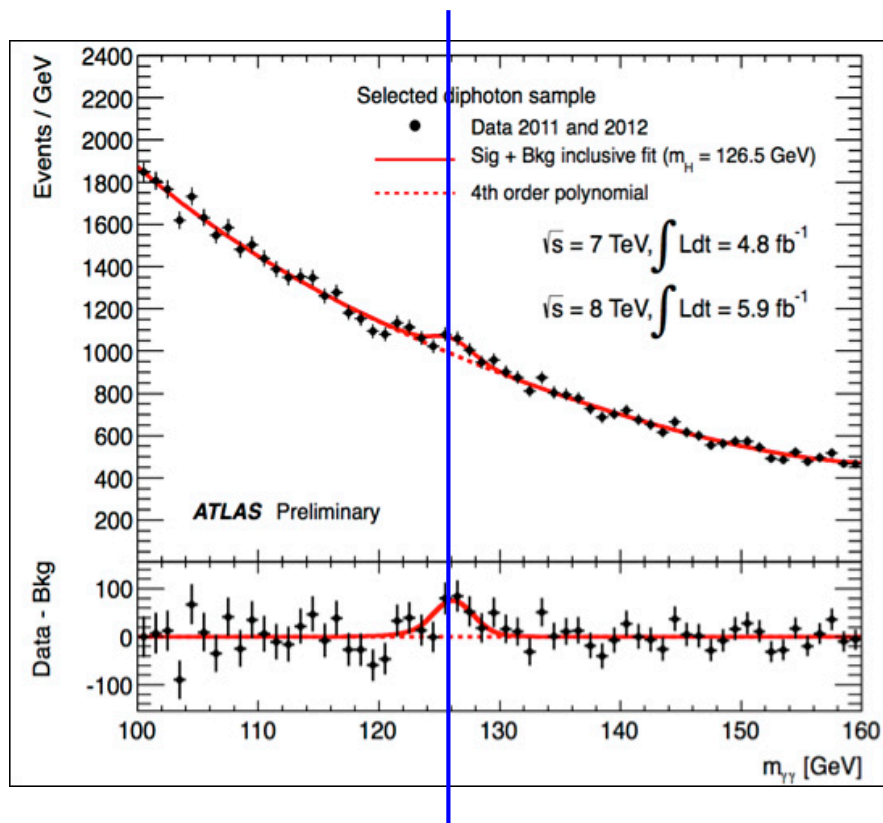
- 10万人の都市で1千人に出口調査
(1%の開票率)

候補者A	候補者B	候補者C	候補者D	候補者E
600	200	100	50	50

候補者Aの票が実際は過半数を割っているのに、
600以上の票が得られる確率は 10^{-10} 以下

仮説検定の使いどころ

Higgs粒子の発見



有意水準 5σ で棄却

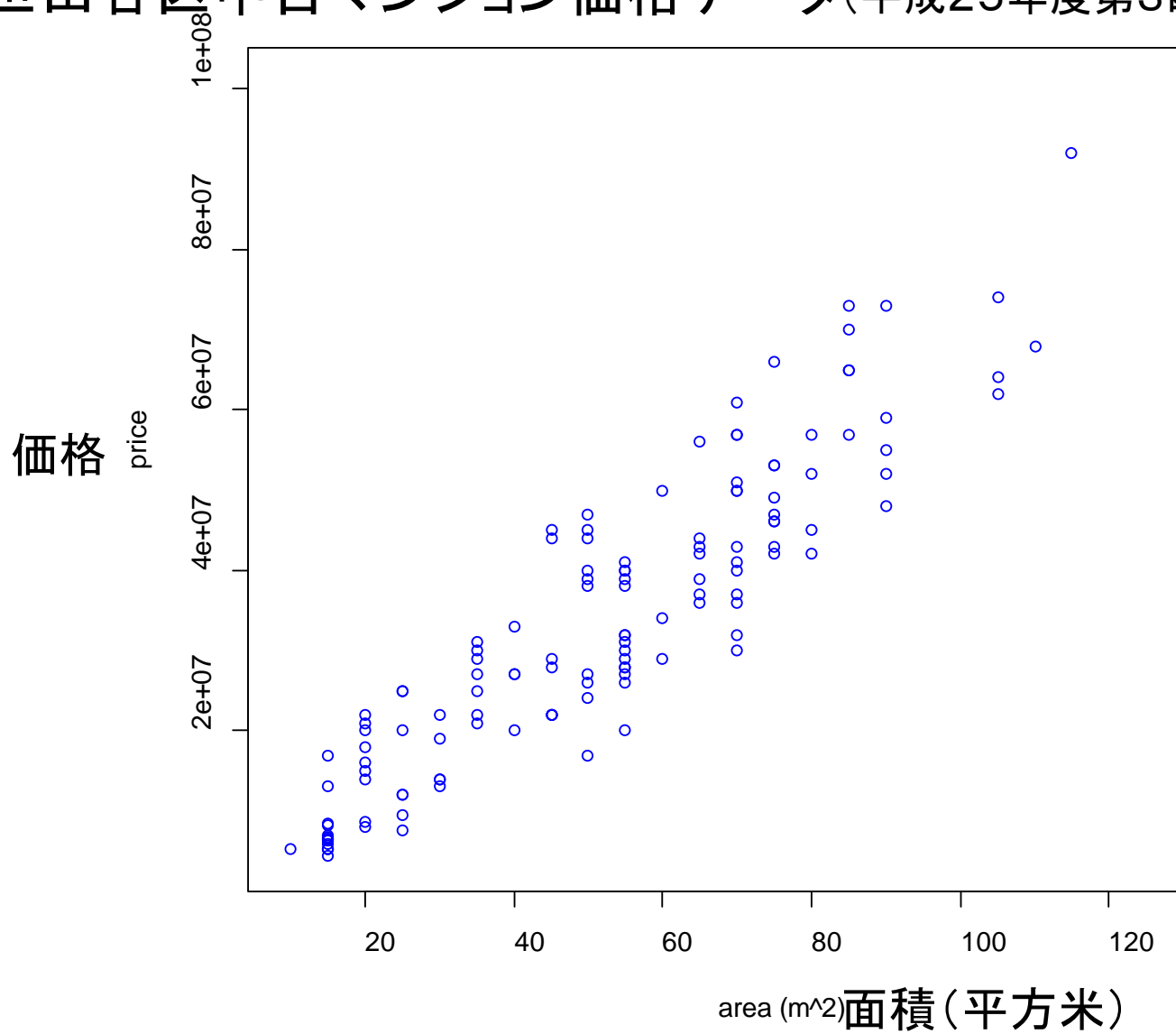
ヒッグス粒子が存在しないと仮定したときにこのような現象が起きる確率は
1/300万以下

統計学の考え方

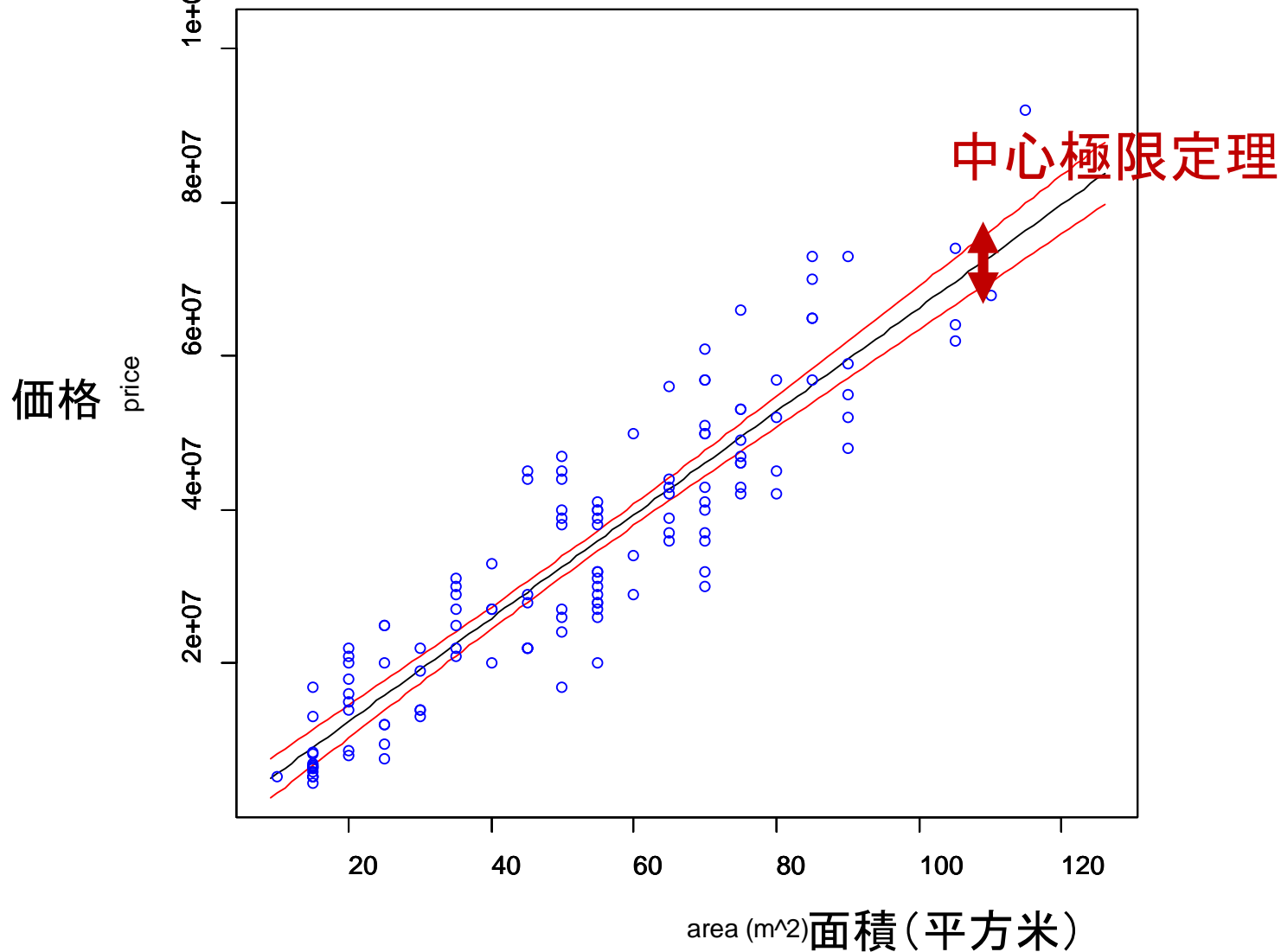
「データを集めると真実が見えてくる」

回帰分析

世田谷区中古マンション価格データ(平成25年度第3四半期分)



世田谷区中古マンション価格データ(平成25年度第3四半期分)



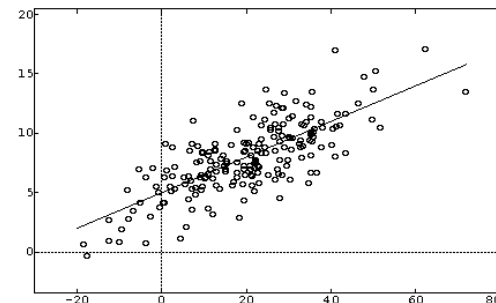
回帰分析の役割

ある確率的な事象を他の変数で説明できる。

- 病気のなりやすさ
- 物の価格

他にも...

- ある疾病にどの遺伝子が関わっているか？
- 大気汚染の要因
- 地震の兆候の発見

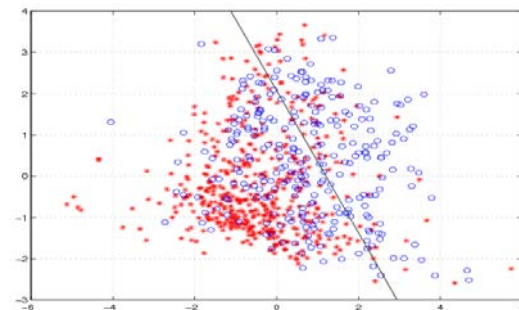


統計学の役割

データ解析の普遍的な方法論を提供

データを用いた法則の発見

- 経験科学(実験)
- 理論科学(理論・数式)
- 計算科学(シミュレーション)
- **データ科学**



統計学の歴史

「統計」という言葉

- Statistics (英語)

ラテン語「status」: 国家・状態

国家の人力・財力といった**国勢データ**

転じてデータを用いた分析全般を指すように



初代ローマ皇帝アウグストゥス: センサス (Census), 人口や土地を調査

確率論

もとは賭博の考察から始まる

ガリレオ:「さいころゲームについての考察」

パスカル&フェルマー:サイコロ賭博の考察から標本理論の概念

ニコラス・ベルヌーイ
二項分布 (1713年)

トーマス・ベイズ (18世紀)
ベイズの定理 (1764年) → ベイズ統計学



ピエール=シモン・ラプラス (18-19世紀)
中心極限定理
ラプラス分布

近代統計学の確立

- ロナルド・フィッシャー (1890~1962)



イギリスの統計学者

近代統計学の枠組みを整備 (1930年代)

最尤推定, 統計的十分性, フィッシャー情報量
分散分析, 実験計画, . . .

もう一つのデータ科学

機械学習

機械学習とは

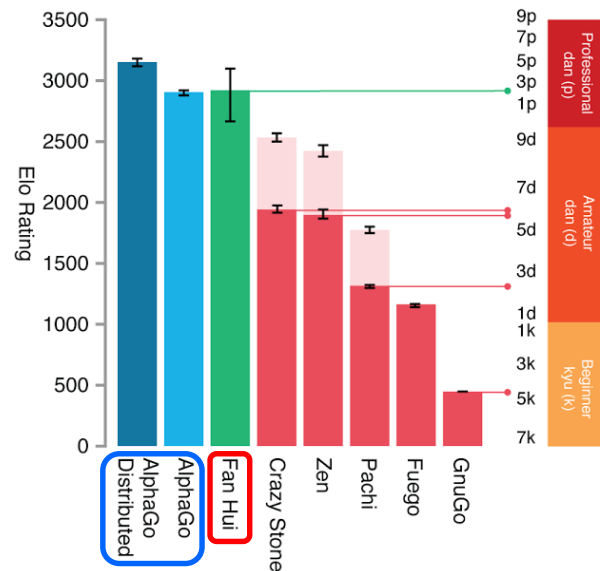
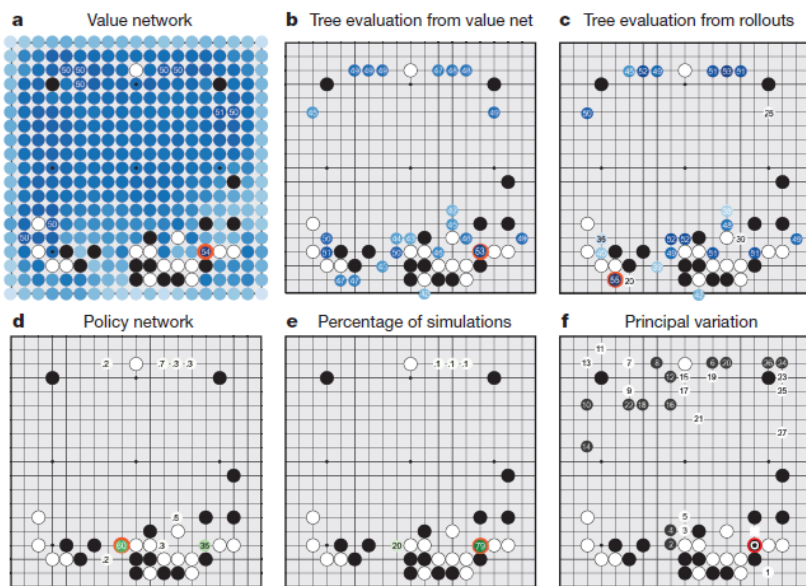
機械が賢くなるための方法論

Wikipedia「人工知能における研究課題の一つで、人間が自然に行っている学習能力と同様の機能をコンピュータで実現しようとする技術・手法のことである。」

Arthur Samuel「Field of study that gives Computers the ability to **learn without being explicitly programmed**」(1959)

AlphaGo

深層学習+強化学習+自己学習



囲碁は将棋よりも難しい問題。

2015年10月5～9日欧州覇者Fan Hui (2段) に5-0で勝利

2016年3月9～15日イ・セドルと対局し4-1で勝利



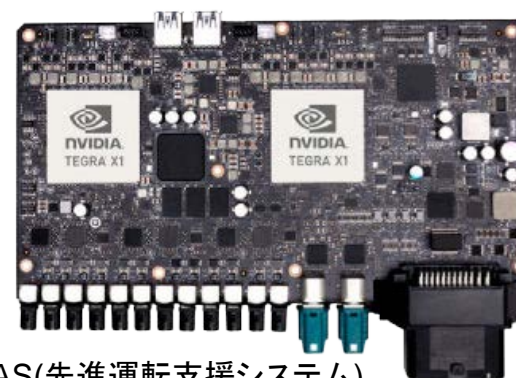
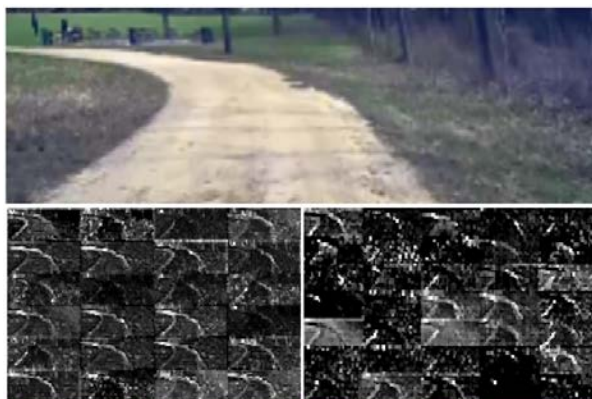
コンピュータ将棋



Bonanza: 2005年6月 **機械学習**の利用
(棋譜データからの評価関数の学習)

第2回将棋電王戦: 2013年3月~4月 3勝1敗1分

自動運転

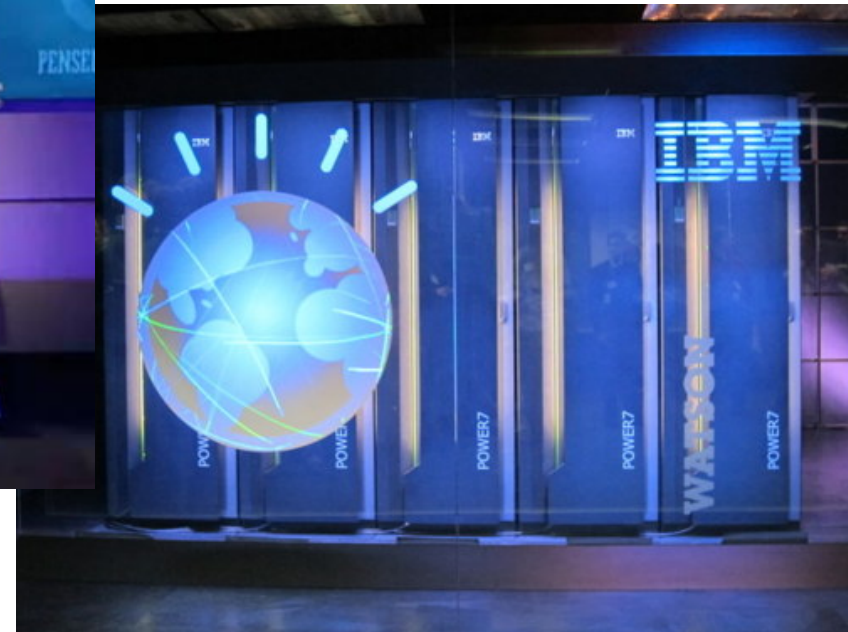


ADAS(先進運転支援システム)

[End to End Learning for Self-Driving Cars, Nvidia 2016]

IBM Watson

クイズ番組「ジェパディ!」 2011年2月14-16日



機械学習・自然言語処理

- 2880個のプロセッサ・コア
- 2億ページ分のテキストデータ
(約100万冊の書籍に相当)

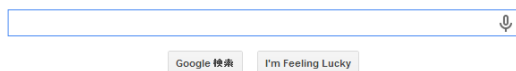
Q: 米国が外交関係を持たない世界の四カ国のうち、最も北にある国

A: 北朝鮮

身近にあふれる機械学習

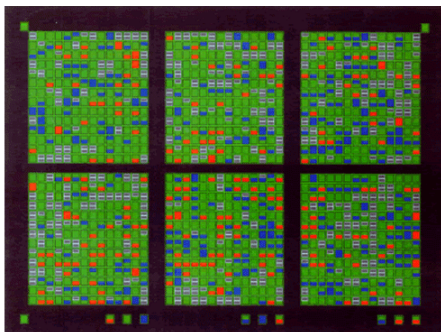
検索エンジン

Google
日本



Google 検索 I'm Feeling Lucky

遺伝子データ解析



音声認識



推薦システム

amazon.co.jp [マイストア](#) | [Amazonポイント](#) | [ギフト券](#) | [タイムセール](#) | [出品サービス](#) | [ヘルプ](#)

カテゴリー [からさがす](#)

[マイストア](#) | [マイページ](#) | [お客様へのおすすめ](#) | [おすすめ商品を正確にする](#) | [プロフィール](#) | [詳しくはこちら](#)

[マイストア](#) > [おすすめ商品](#)
もしあなたが さんではない場合、[サインイン](#)してください

本日のみ
[おすすめ商品を見る](#)

おすすめ商品
[Amazonインスタントビデオ](#)
[Amazonファッション](#)
[Android アプリストア](#)
[DIY・工具](#)
[DVD](#)
[Kindleストア](#)
[MP3ストア](#)
[PCソフト](#)
[TVゲーム](#)

これらのおすすめ商品は、[顔にお持ちの商品](#)などに基ついています。

表示: [すべて](#) | [ニュースリリース情報](#) | [まもなく発売](#)

16.  **トワイニング クオリティアールグレイ 100g**
トワイニング
おすすめ度: ★★★★★ (12)
在庫あり
価格: ¥ 864
新品の出品: 6月 4日より

持っています 興味がありません ★★★★★ この商品を購入する
モンテニス ティーマシジャー ゴールドなどを購入されたお客様におすすめします (おすすめの商品に反映させる商品)

機械学習プラットフォーム

 Google Cloud Platform

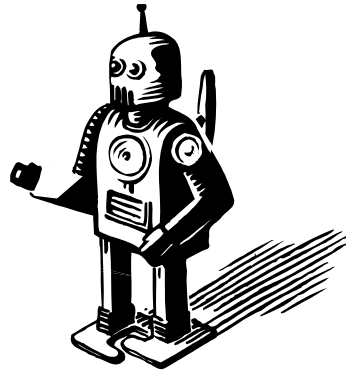
 Microsoft Azure

 amazon
web services™

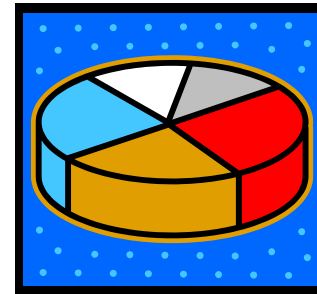
機械学習と統計学

機械学習と統計学には深い関係がある。

機械



統計



データからの情報抽出

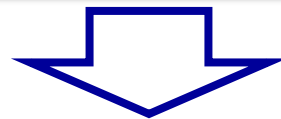
乱雑なデータからその裏にある規則を見つける

機械学習の考え方

将棋の強いプログラムを作ろう！

強い手を指すようにプログラムする？

→ プログラマーが将棋に強くないといけない
そもそも盤面のパターンは 10^{220} 通り以上
無理...



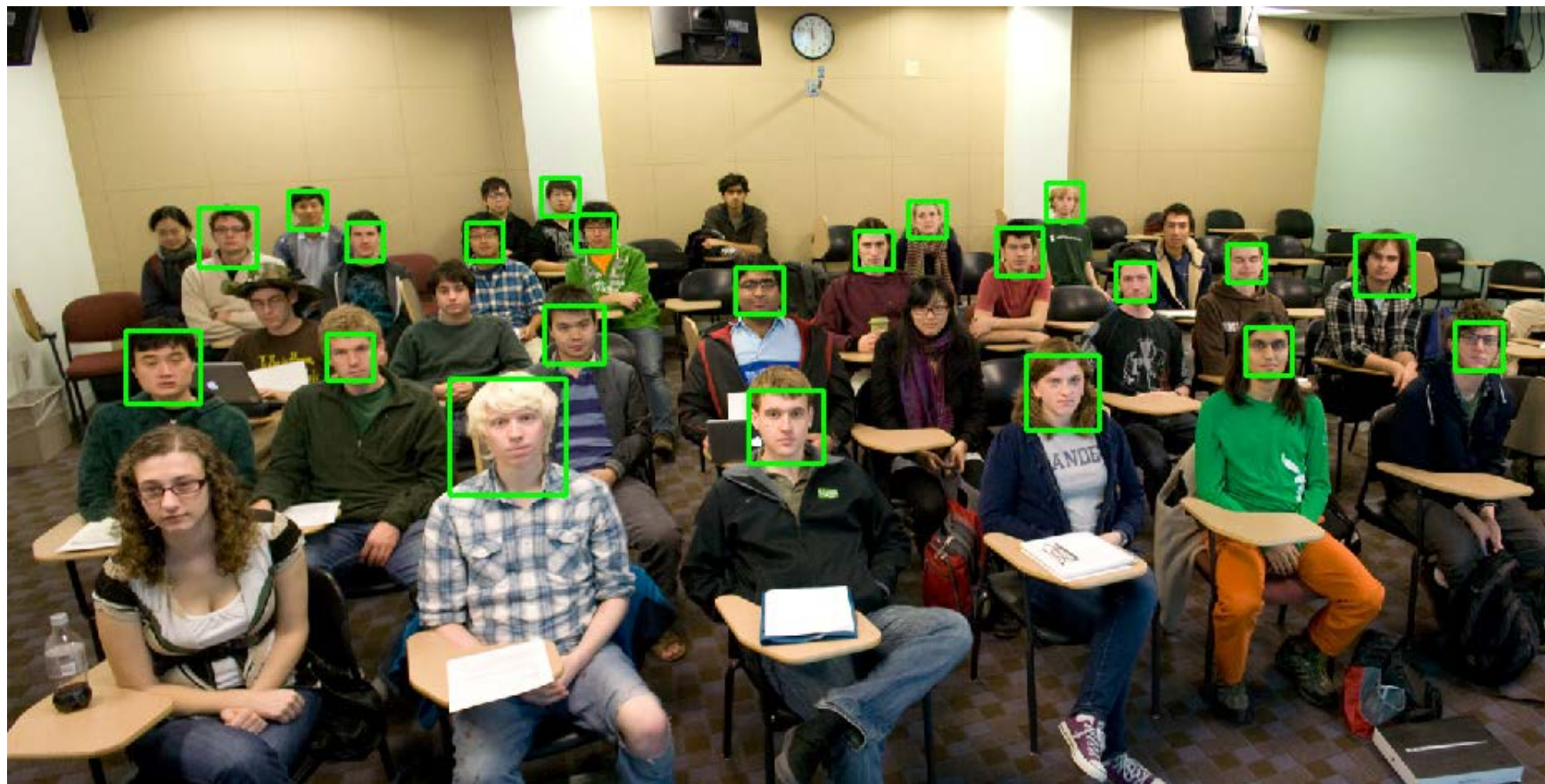
具体的にどうやって指すかはプログラムしない！

プログラムするのは**学習の仕方**。

機械にデータから学習させる。

統計学の出番

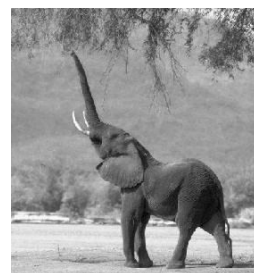
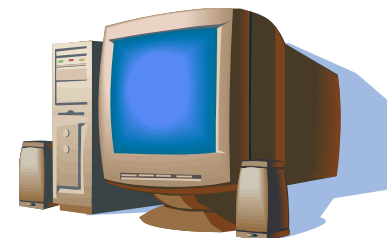
顔認識



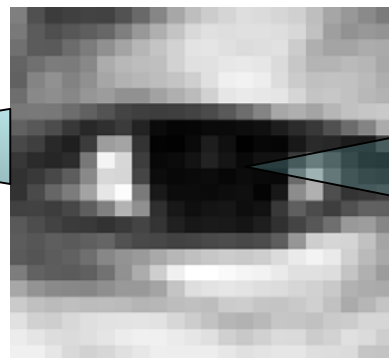
顔認識の学習データ

2001年の方法(今はデータのサイズがもっと大きい)

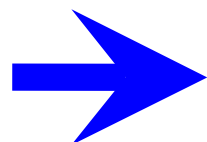
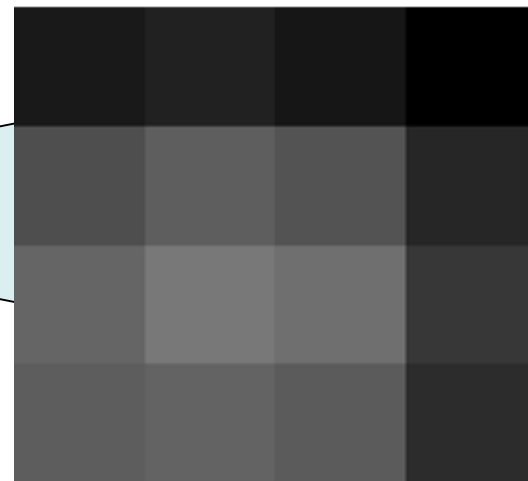
- 約5000枚の顔写真
- 約9500枚の非顔写真



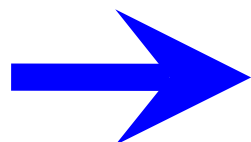
[Viola&Jones,2001]



ピクセル



一列に
並べる



濃淡値

$$\begin{bmatrix} 240 \\ 130 \\ 100 \\ 110 \\ \vdots \\ 220 \end{bmatrix}$$

画像を数字の列で表現
数学的には...

ベクトル

(数万次元)

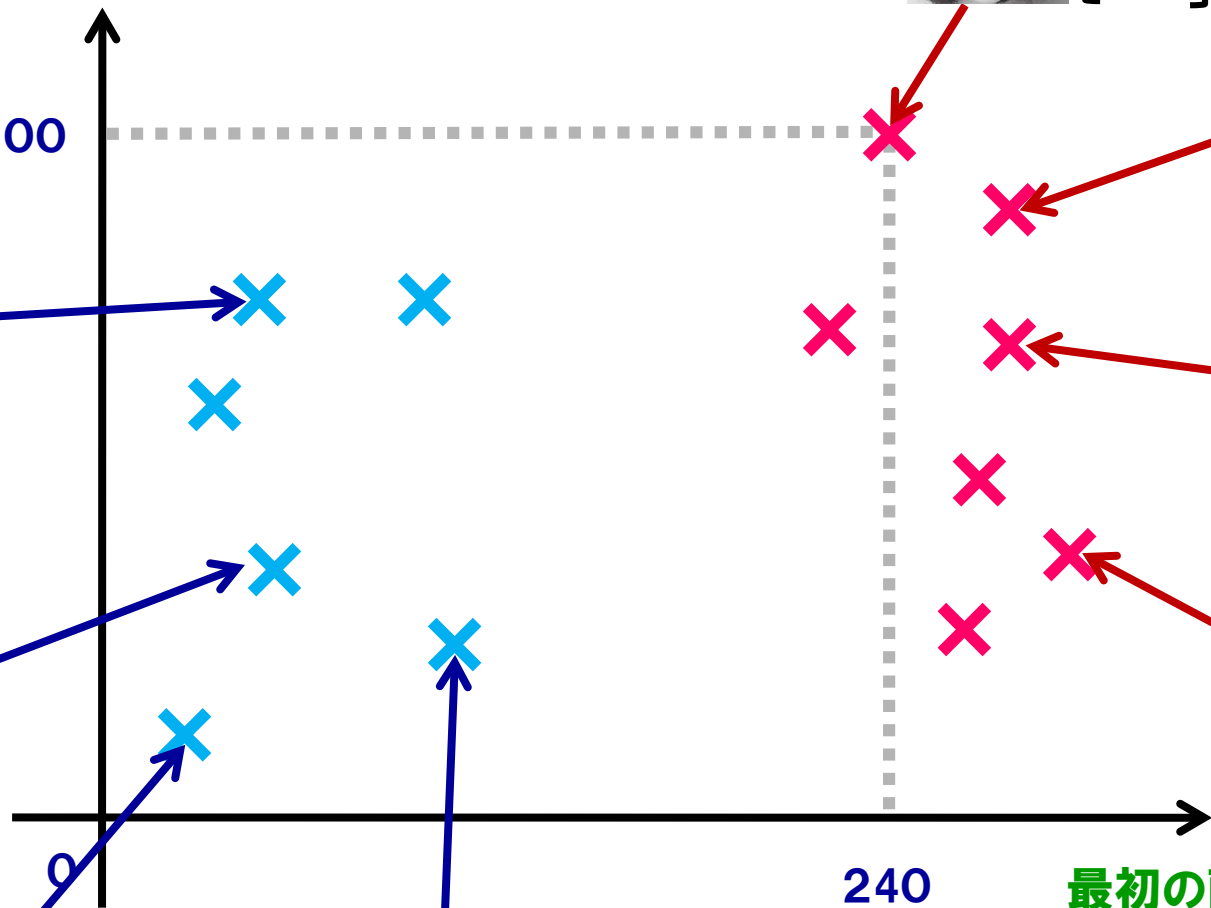
二番目の画素の濃さ

200



最初の画素の濃さ

240



機械に学習させる?

二番目の画素の濃さ

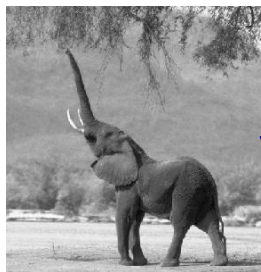
顔じゃない

顔

なるべく遠く

最初の画素の濃さ

(超平面)



機械に学習させる?

二番目の画素の濃さ

顔じゃない

顔

なるべく遠く

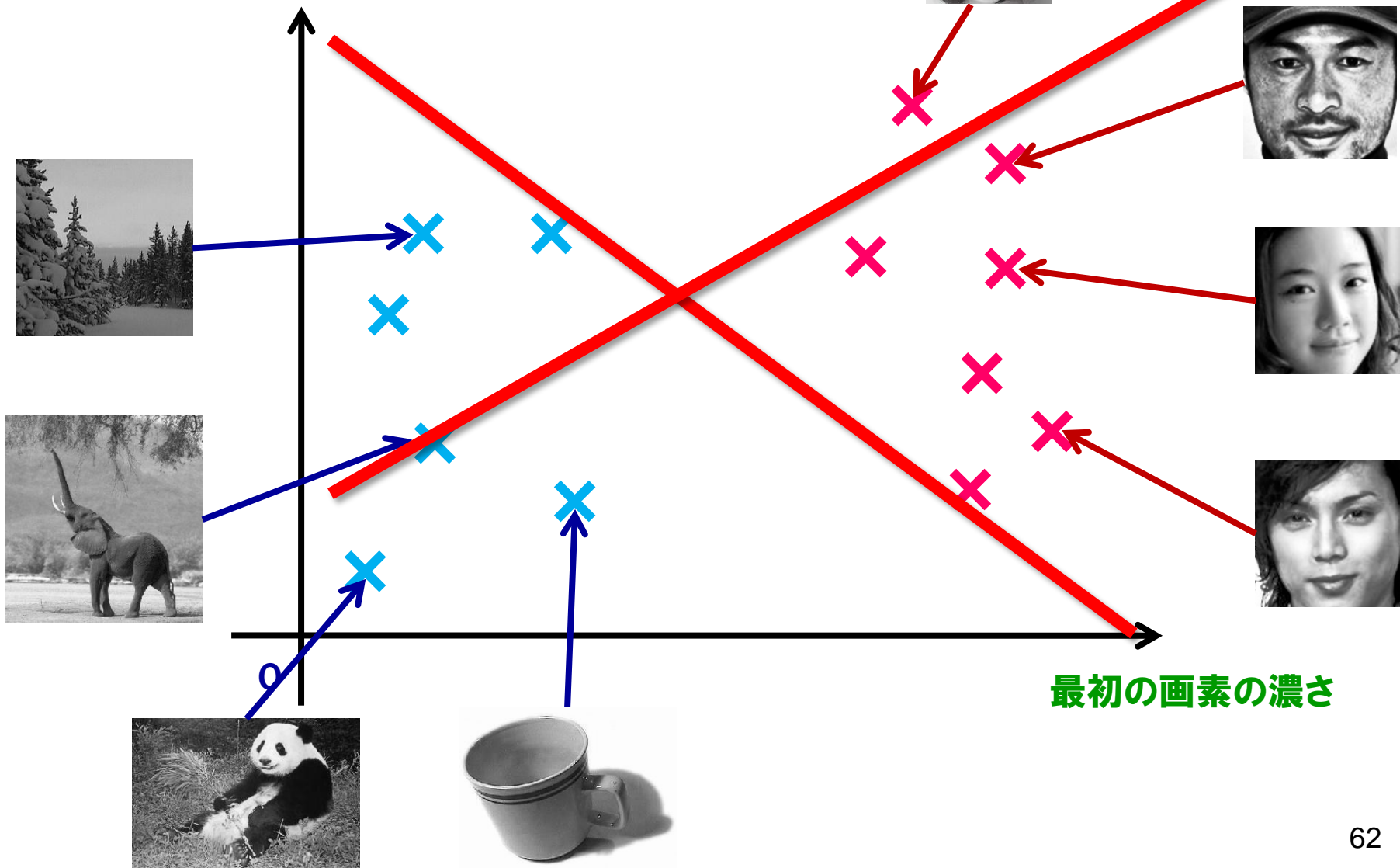
☆見たことのない画像も判別できる。



(超平面)

ダメな例

二番目の画素の濃さ

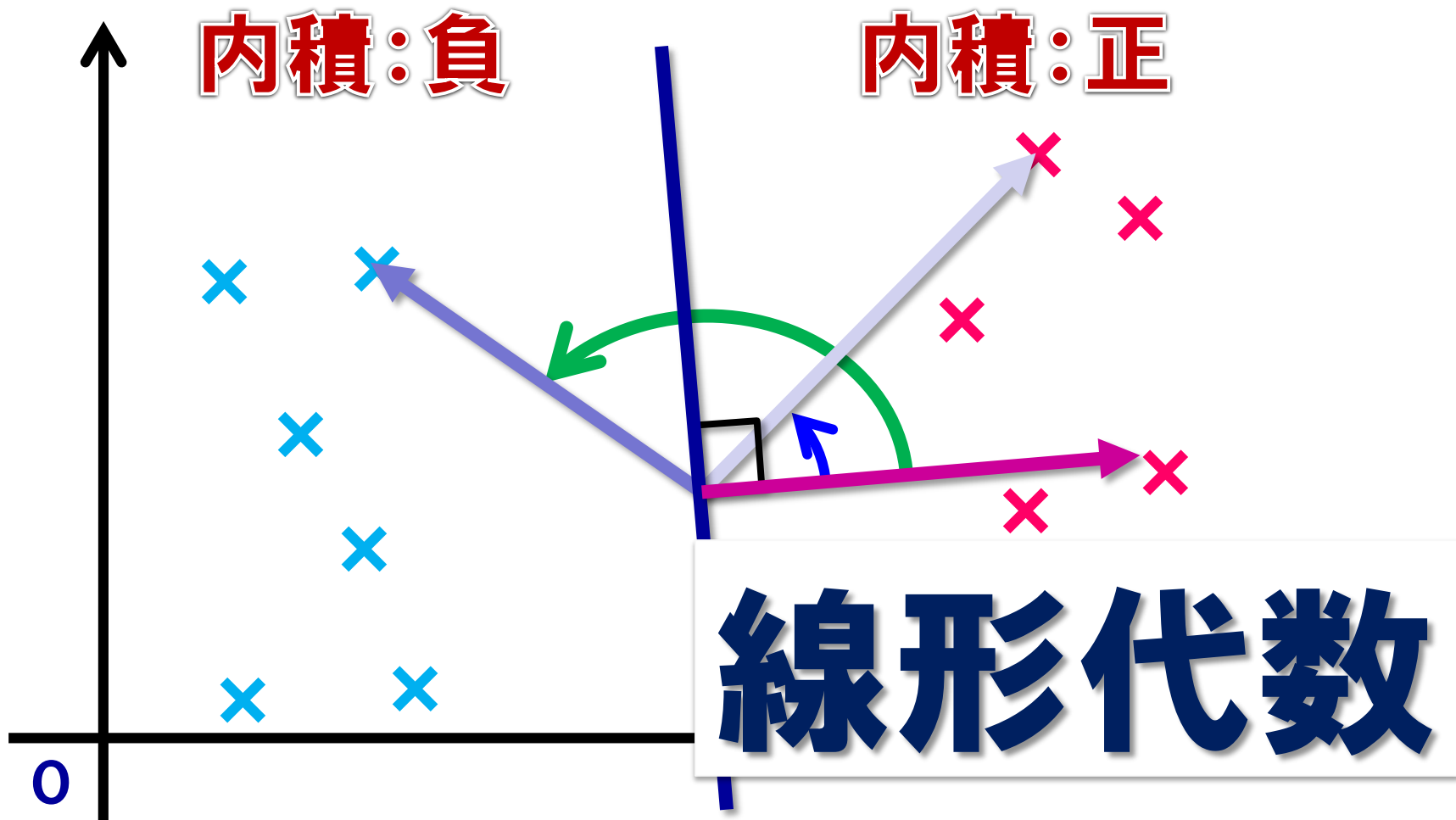


最初の画素の濃さ

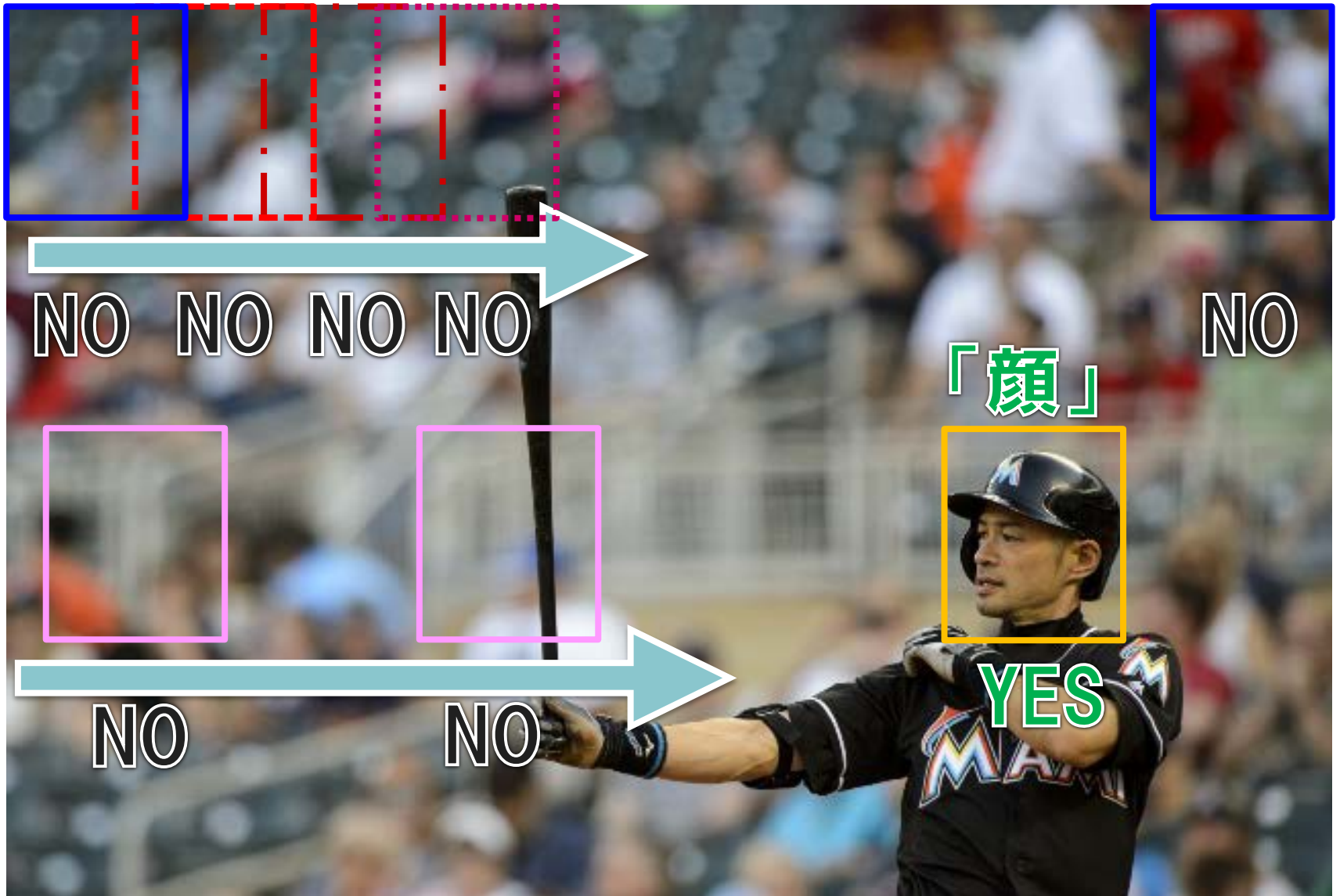
(小話) 「内積」について

超平面の右, 左はどうやって判定する?

→ **内積**をつかう.



顔検出

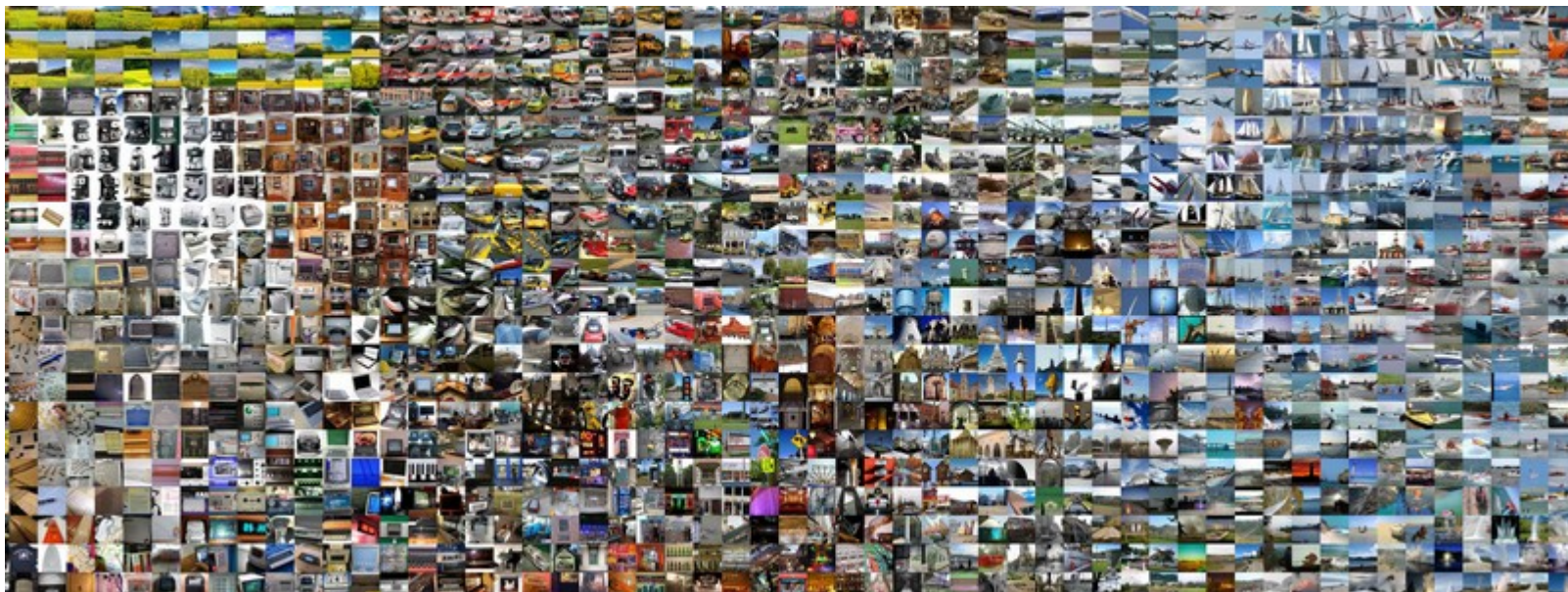


左端から少しずつ領域をずらして顔を探す.

最新技術の紹介 「深層学習」

仕組みは難しいので理解できなくてよいです。
大学で勉強すれば理解できるようになります。

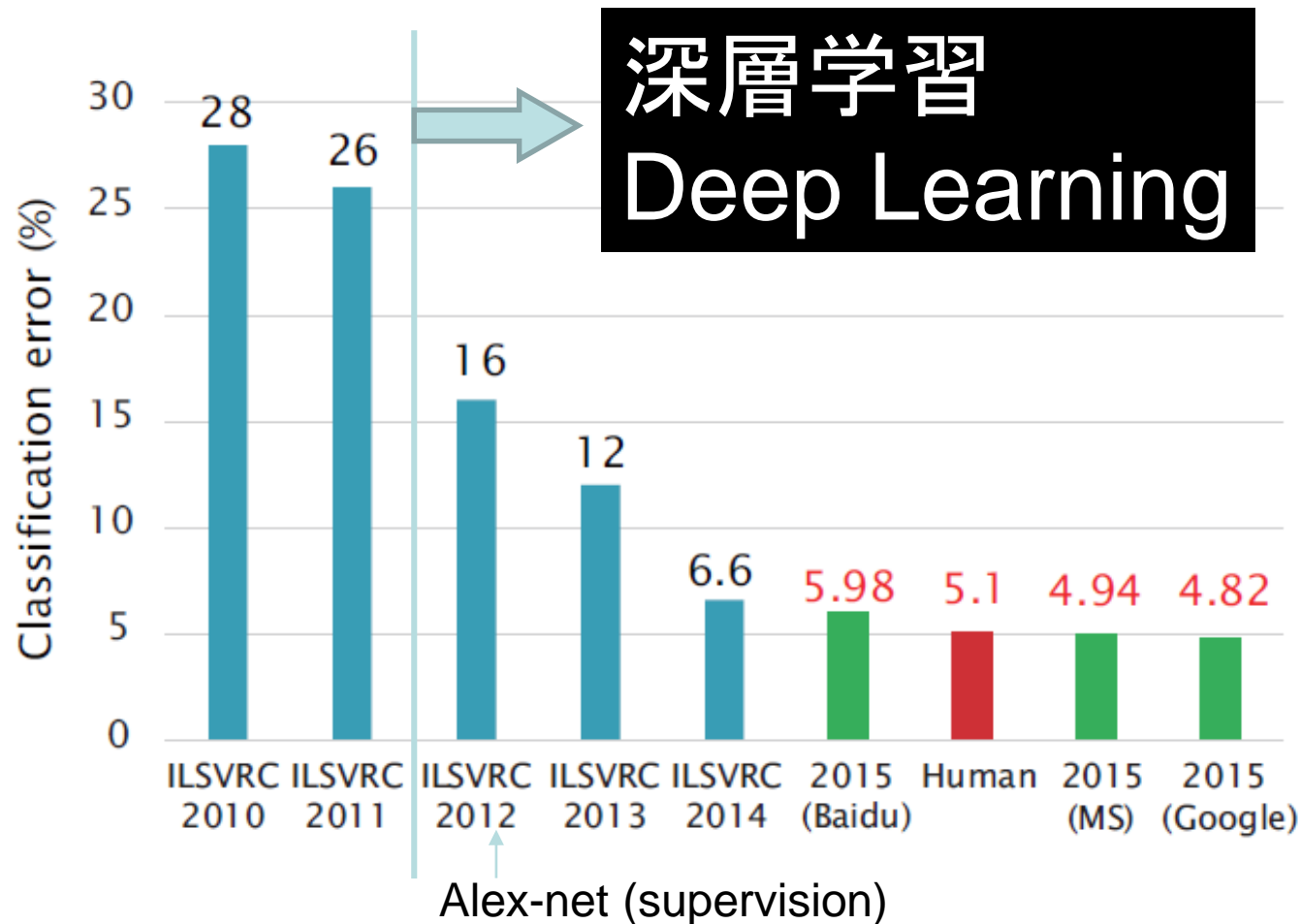
ImageNet



ImageNet: 21841クラス, 14,197,122枚の訓練画像データ
研究用に公開されている

[J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei.
ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.]

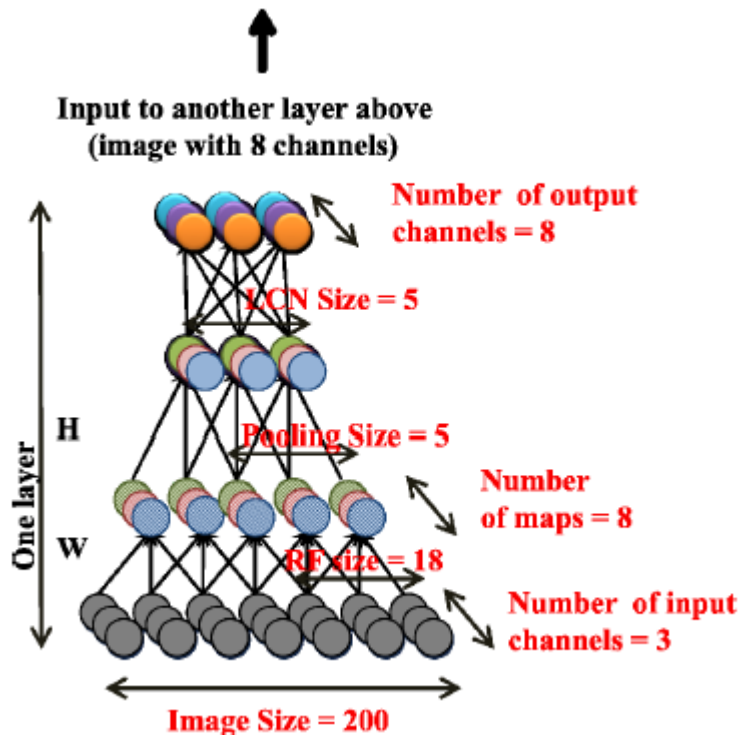
ImageNetデータにおける識別精度の変遷



ImageNet: 21841クラス, 14,197,122枚の訓練画像データ

(図は[中山. 深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習", 電子情報通信学会音声研究会7月研究会, 2015.]より)

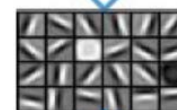
深層学習 (多層ニューラルネット)



3rd layer
"Objects"



2nd layer
"Object parts"



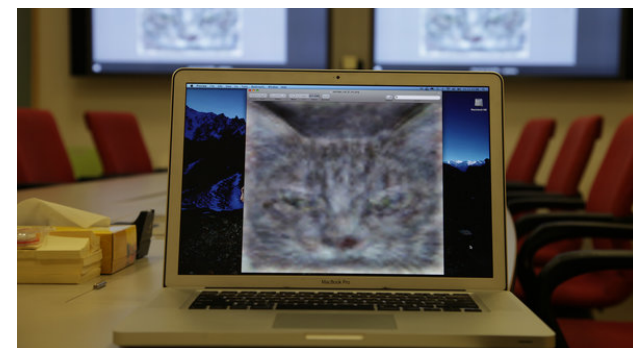
1st layer
"Edges"



Pixels

Youtubeから1000万画像
16,000CPU
一週間計算

[Le et al.: ICML2012]

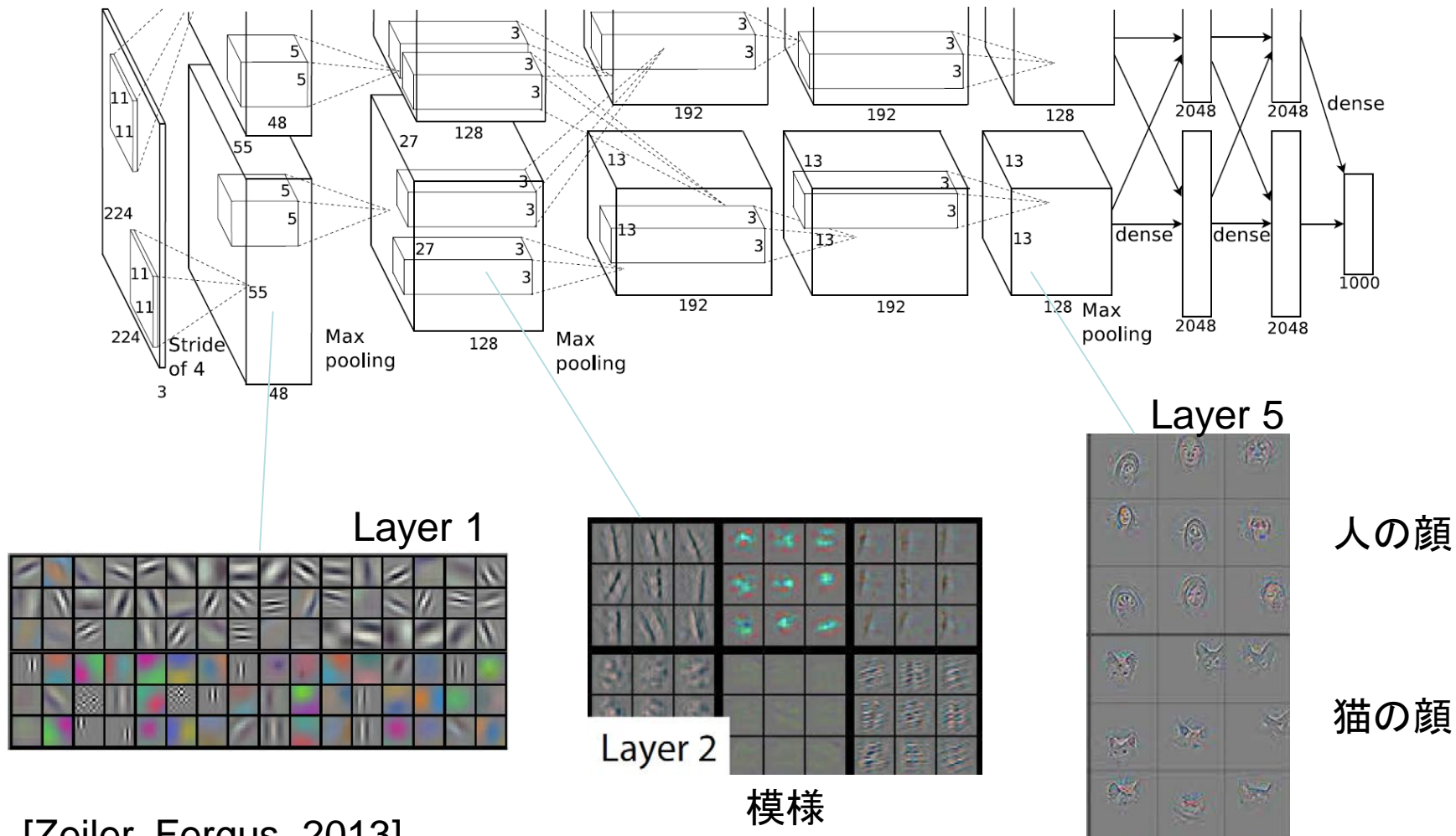


今はここまでしなくても、より高い精度が出せるようになっている。

Alex-net [Krizhevsky, Sutskever + Hinton, 2012]

[Krizhevsky, Sutskever, and Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. pp. 1097--1105, 2012.]

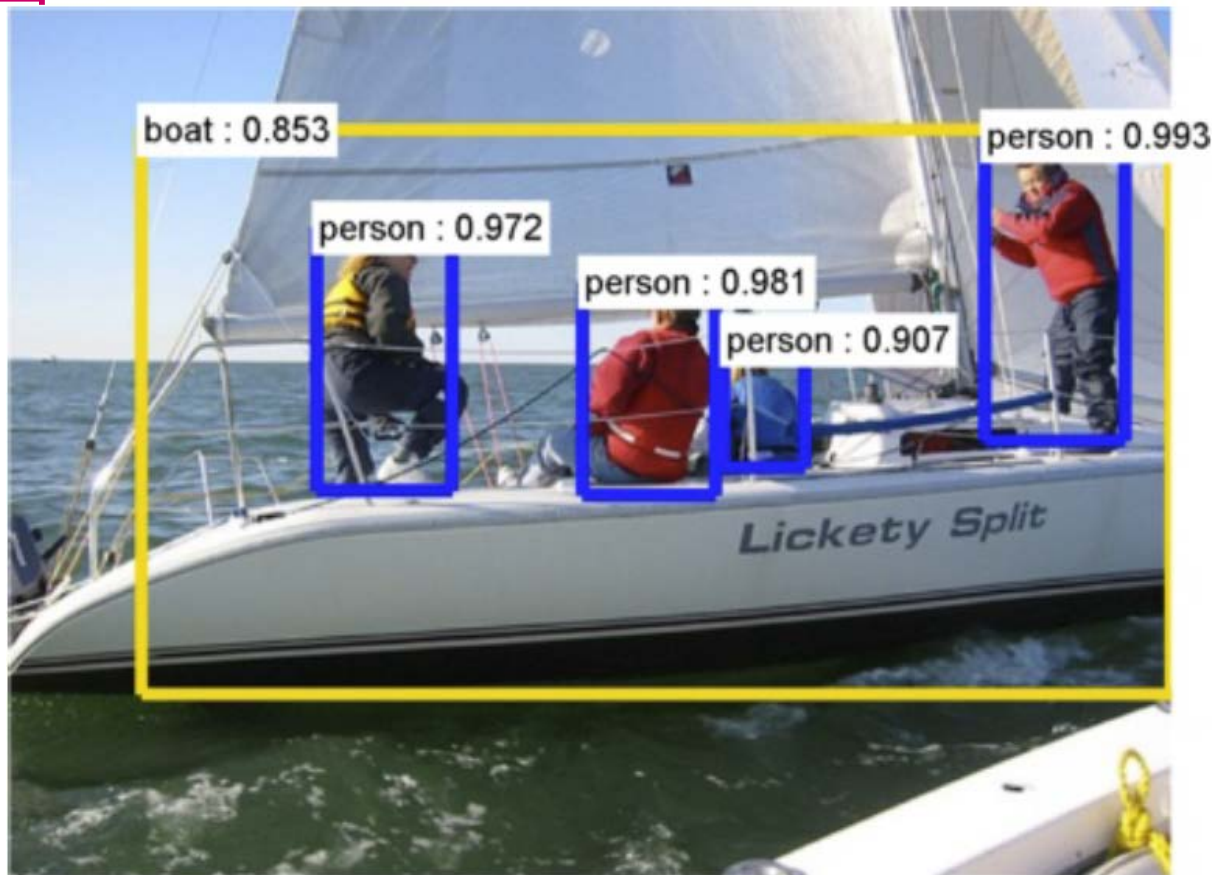
畳み込みニューラルネットを5層積み重ね(+pooling+3層の全結合層)
中間層ではより抽象的な情報がコードされる



[Zeiler, Fergus, 2013]

深層学習の応用

物体検出









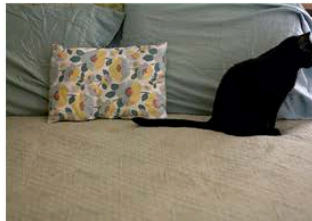




「どこに」+「なにが」
写っているか

キャプション生成

入力画像

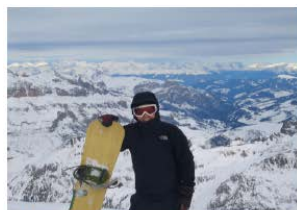
自動生成された説明文

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 <p>A person riding a motorcycle on a dirt road.</p>	 <p>Two dogs play in the grass.</p>	 <p>A skateboarder does a trick on a ramp.</p>	 <p>A dog is jumping to catch a frisbee.</p>
 <p>A group of young people playing a game of frisbee.</p>	 <p>Two hockey players are fighting over the puck.</p>	 <p>A little girl in a pink hat is blowing bubbles.</p>	 <p>A refrigerator filled with lots of food and drinks.</p>
 <p>A herd of elephants walking across a dry grass field.</p>	 <p>A close up of a cat laying on a couch.</p>	 <p>A red motorcycle parked on the side of the road.</p>	 <p>A yellow school bus parked in a parking lot.</p>

Google による画像説明文章の自動生成

[Vinyals et al., Show and tell: A neural image caption generator. CVPR, 2015]

Visual Question Answering



Q: what is the man holding a snowboard on top of a snow covered? A: **mountain**



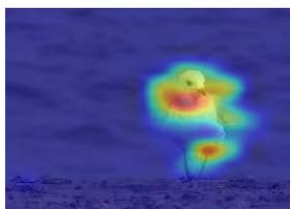
Q: what is the color of the bird? A: **white**



Q: how many snowboarders in formation in the snow, four is sitting? A: **5**



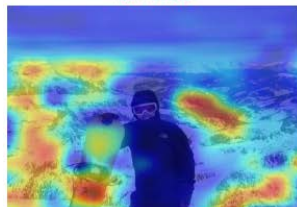
what is the man holding a snowboard on top of a snow covered



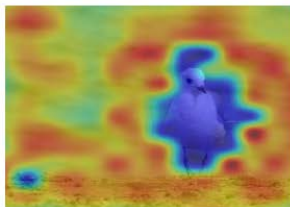
what is the color of the bird ?



how many snowboarders in formation in the snow , four is sitting ?



what is the man holding a snowboard on top of a snow covered ?



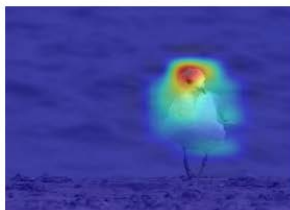
what is the color of the bird ?



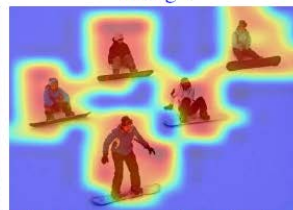
how many snowboarders in formation in the snow , four is sitting ?



what is the man holding a snowboard on top of a snow covered ?



what is the color of the bird ?



how many snowboarders in formation in the snow , four is sitting ?

Q: 鳥の色は何色か？ A: 白

画像の内容に関する質問に答える。

画像のスタイル変換

元画像



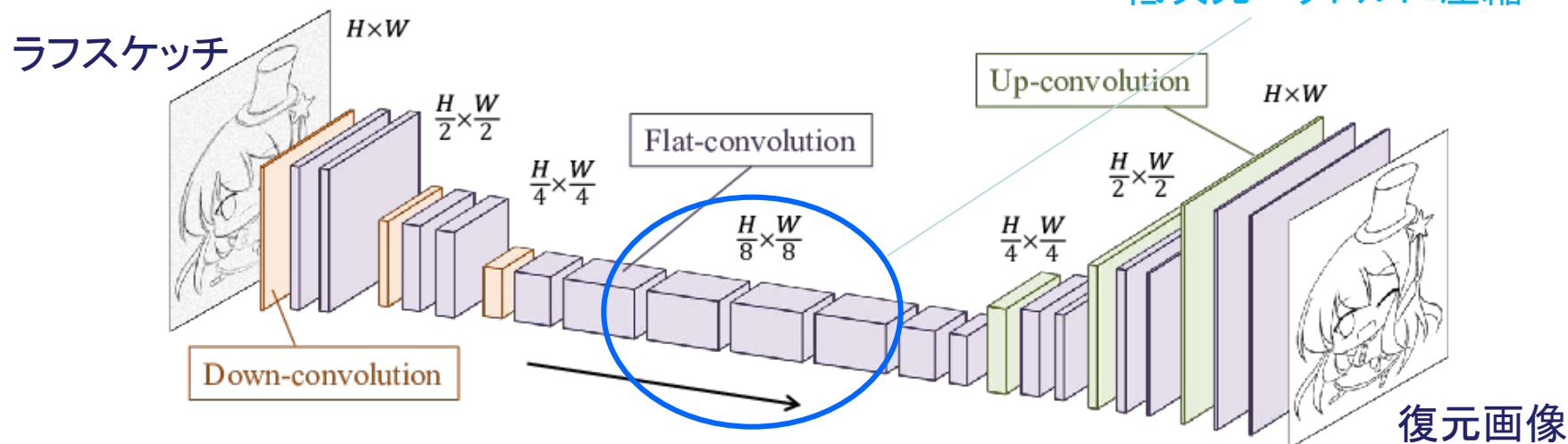
参照画像



参照画像のような画風に変換

ラフスケッチの自動線画化

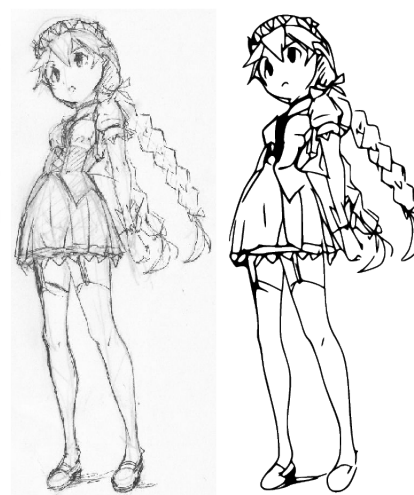
低次元ベクトルに圧縮



(c) Masks



(d) Book



(e) Standing girl

いろいろな企業で活用

- ・ アメリカ: Google, Facebook, Baidu, . . .
- ・ 日本: トヨタ, リクルート, PFI, Fanuc, . . .



FANUC



- ・ 音声認識, 自然言語処理, 自動運転, ウェブデータ解析, などなど
- ・ 数学が直に役に立っている.
- ・ アメリカでは機械学習・人工知能が専門の博士号取得者が高額報酬で多数雇用されている.

最後に

- 長い統計学の歴史
 - 17世紀: 国勢調査
 - 1930年代: フィッシャーによる統計学の整備
- 現代のニーズ
 - 2012年: 深層学習が広がり始める
 - 2015年~: 人工知能ブーム到来

**数十年後の世界
何が重要か？**

中心極限定理の意味

- **どんな分布**であろうとも、たくさんサンプルを取ってきて平均を取れば、**すべて同じ分布**（正規分布）に従う。

普遍性

→ 偶然が支配する世界に隠れた規則性。
一般的な方法論が構築できる。

統計学の成立

「コイン投げが科学になる」