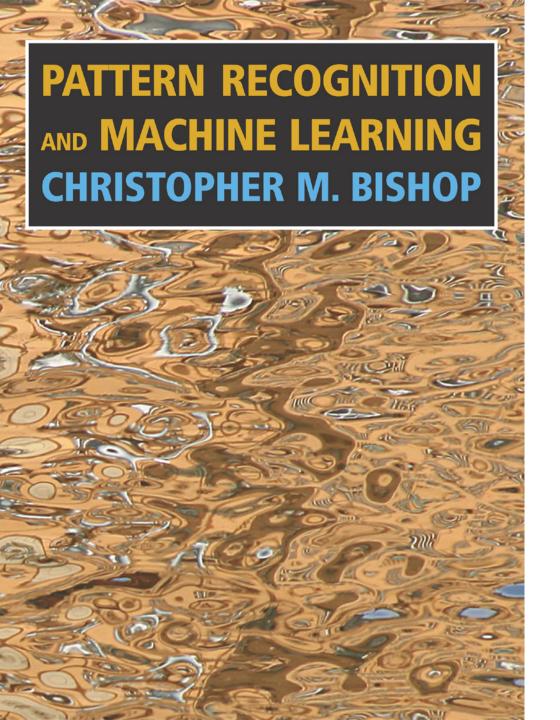
## introduction

正田 備也

masada@rikkyo.ac.jp



## 参考書

C. Bishop. *Pattern Recognition and Machine Learning*. 2006.

- この授業はベイズ的な統計モデリングの授業です
- 他にもいろいろな文献を参考にしつつ内容を構成しています

#### 今日のお題

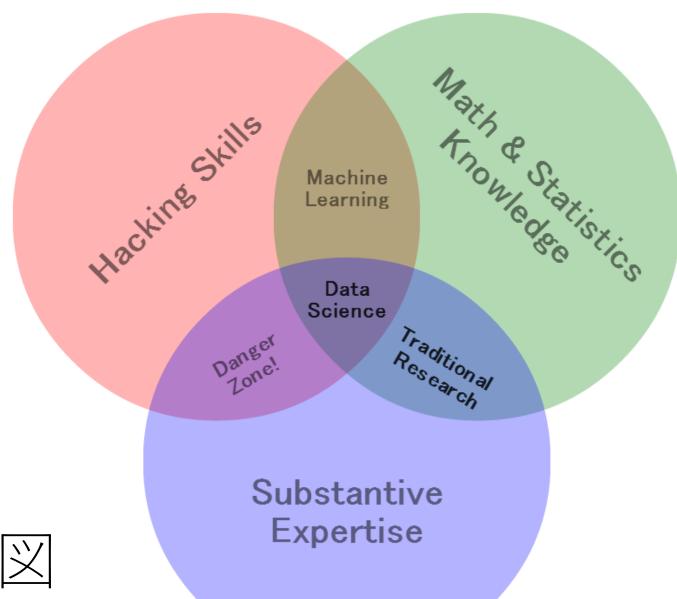
- 統計モデリングとは
  - unsupervised / supervised
  - generative / discriminative
- ・確率の復習
  - ・確率変数、同時確率、周辺化、条件付き確率、ベイ ズ則、確率分布、期待値・・・

# 統計モデリングとは

## 科学全体の中でのデータ科学の位置づけ

https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/

	計算機なし	計算機あり
演繹的	理論科学	シミュレーション
帰納的	実験科学	データ科学



データ科学のベン図

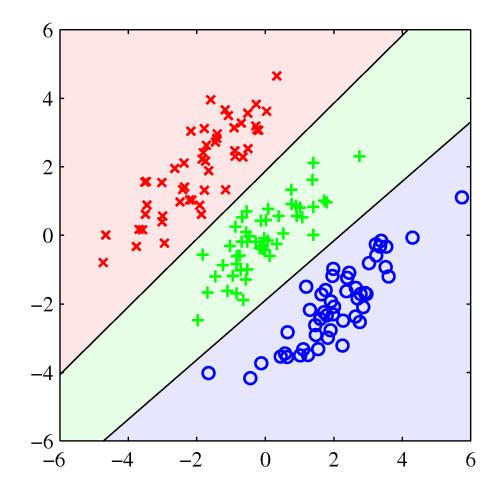
and the second

https://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

#### 機械学習の2区分 (既習)

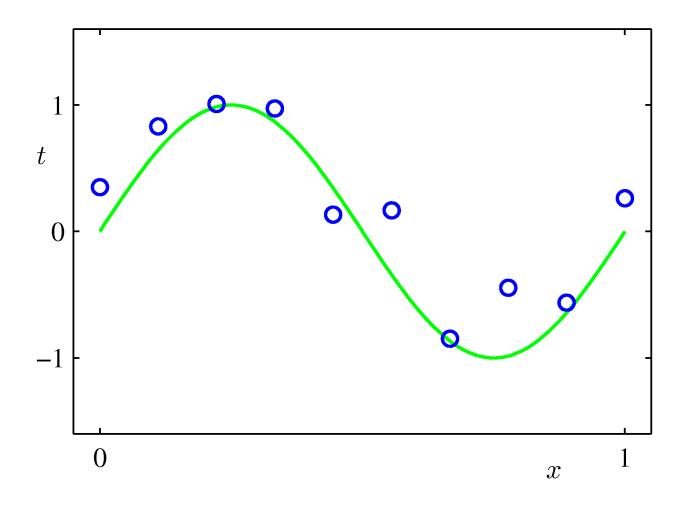
- 教師あり学習 (supervised learning)
  - classification
  - regression
- 教師なし学習 (unsupervised learning)
  - clustering
  - dimensionality reduction

#### classificationのイメージ



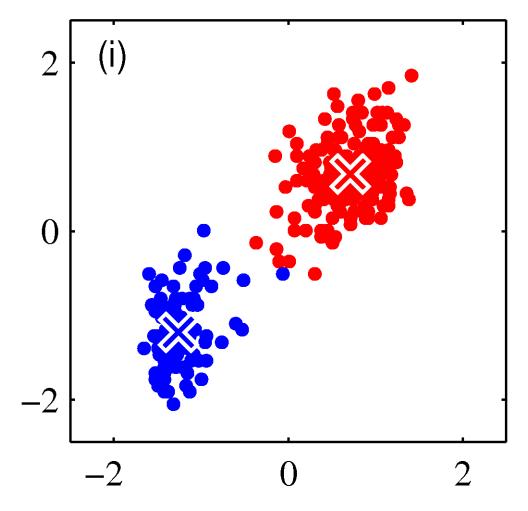
C. Bishop. Pattern Recognition and Machine Learning. 2006.

## regressionのイメージ



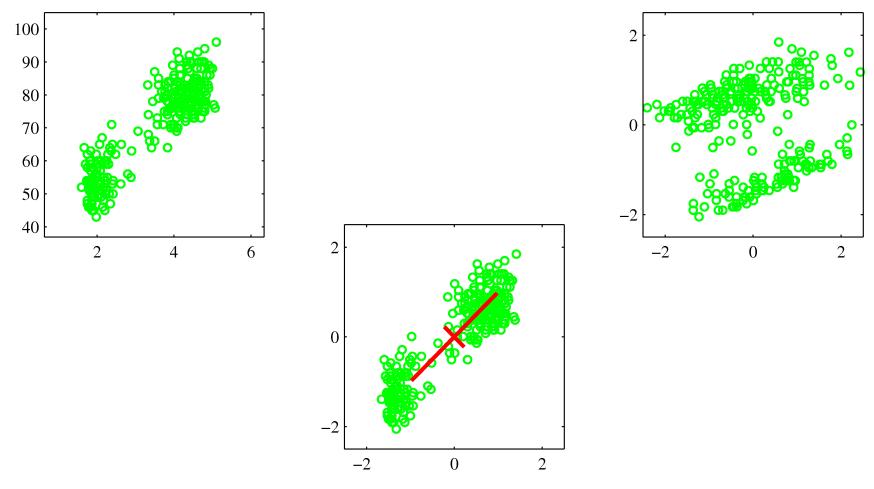
C. Bishop. Pattern Recognition and Machine Learning. 2006.

## clusteringのイメージ



C. Bishop. Pattern Recognition and Machine Learning. 2006.

#### dimensionality reductionのイメージ

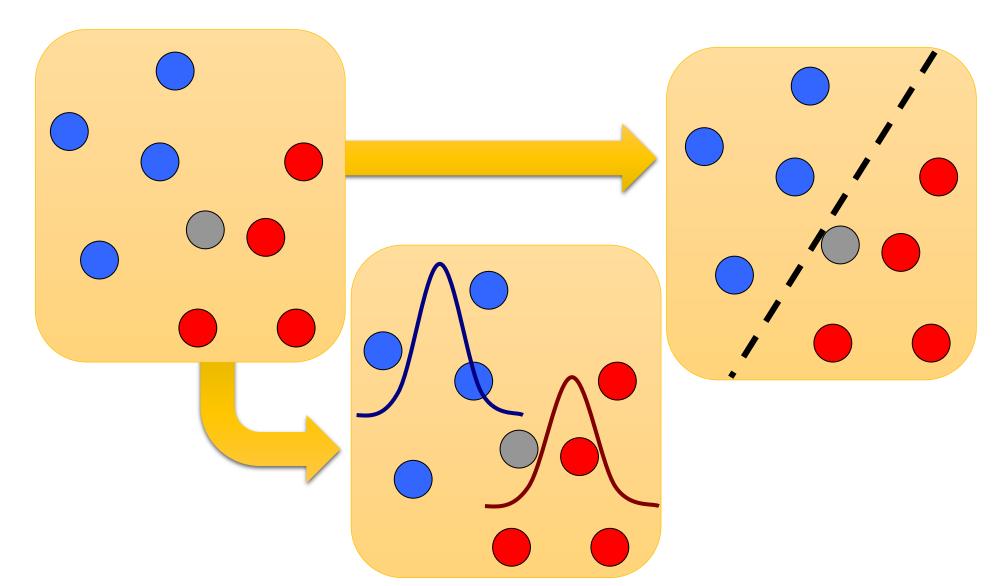


C. Bishop. Pattern Recognition and Machine Learning. 2006.

## 機械学習の別の2区分

- クラスを識別する境界を見つける
  - 境界から遠いところにあるデータは気にしない
- クラスを生成する分布を見つける
  - 境界はこの分布 distribution の違いから派生する

## boundary vs distribution



#### discriminative vs generative

- discriminative methods
  - クラスを識別する境界を計算機に学習させる
- generative methods
  - クラスを生成する確率分布を計算機に学習させる
    - この授業ではこちらを教えます。

## generative approaches (1/2)

'Generative models in machine learning posit that there is some underlying (…) process that is generating the data you are observing and aim to use the data to <u>infer the parameters</u> of that underlying process, which then lets you classify the data.'

http://www.forbes.com/sites/quora/2015/02/12/what-is-the-future-of-machine-learning/

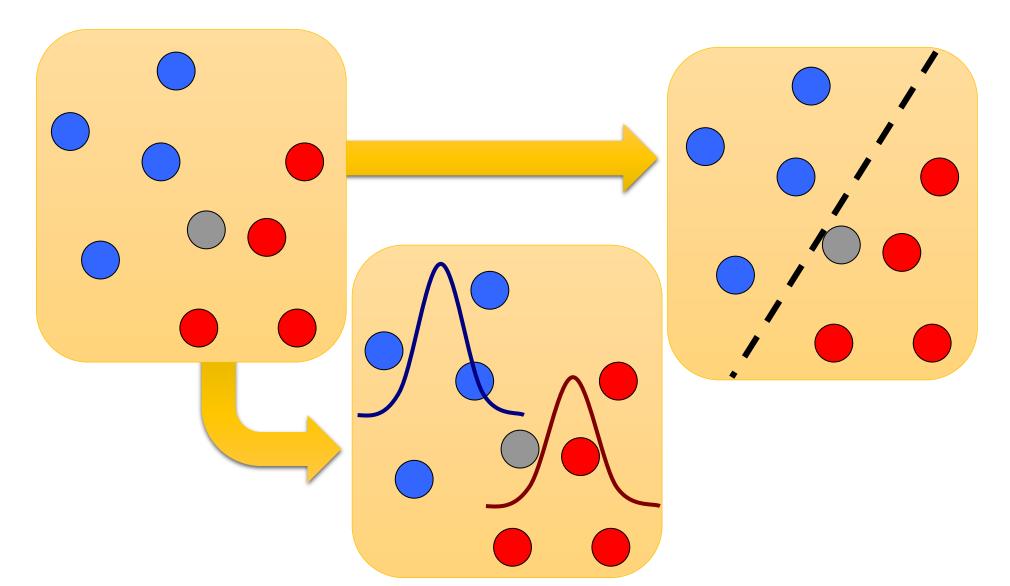
## generative approaches (2/2)

'In my mind, if you succeed in solving a generative model, you have "understood" the data and the problem.'

'Unfortunately for me, discriminative models tend to work better than generative models to solve lots of machine learning problems'

http://www.forbes.com/sites/quora/2015/02/12/what-is-the-future-of-machine-learning/

## boundary vs distribution



データを生成する確率分布を推定することの難しさ http://ibisml.org/archive/ibisml001/Sugiyama.pdf

「統計的機械学習のほとんどの課題は、データの生成確率分布の推定を介して解決することができる。

しかし、確率分布の推定は機械学習における最も困難な問題の一つとして知られているため、現実的には分布推定を回避しながら対象となる課題を解決することが望ましい。」

## GAN [Goodfellow+ 14]

• 観測データの生成確率分布を推定する手法

- 1. 観測データの分布とモデルの出力の分布の密度比を推定する
  - 正規化されていないので分布の形しか分からないため
  - データがどちらから来たかを予測する識別モデルを学習
    - [Goodfellow+ 14]のProposition 1
- 2. 推定された密度比を1に近づける
  - [Goodfellow+ 14] Ø Theorem 1

# 確率の復習

#### 確率の復習

- 確率変数 random probability
- 同時確率 joint probability
- 周辺化 marginalization
- 条件付き確率 conditional probability
- ベイズ則 Bayes rule
- 確率分布 probability distribution
- •期待值 expectation

#### 確率変数 random variable

- •確率変数:どの値をとるかが不確かuncertainな変数
- •p(x=v)は、確率変数xが値vをとる確率、という意味
- $\sum_{x} p(x) = 1$ を満たす
  - つまり、xがとりうる値を $v_1, v_2, ..., v_W$ とすると、以下を満たす $p(x=v_1)+p(x=v_2)+...+p(x=v_W)=1$ 
    - このように、「 $\sum_{x} p(x) = 1$ 」という書き方は、省略を含んでいる
      - こういうのにも慣れましょう

## 同時確率 joint probability

$$p(x = v, z = c)$$

- •上は、確率変数xが値vをとり、かつ、確率変数zが値cをとる確率
- $\sum_{x}\sum_{z}p(x,z)=1$ を満たす。つまり・・・
- ・確率変数zがとりうる値を $c_1, c_2, ..., c_K$ とすると、以下が成り立つ  $p(x=v_1, z=c_1) + \cdots + p(x=v_1, z=c_K) + p(x=v_2, z=c_1) + \cdots + p(x=v_2, z=c_K) + \cdots + p(x=v_W, z=c_1) + \cdots + p(x=v_W, z=c_K) = 1$

## 周辺化 marginalization

$$p(x) = \sum_{z} p(x, z)$$

•p(x)はp(x,z)から上の計算で得ることができる。つまり・・・

$$p(x = v) = p(x = v, z = c_1) + \dots + p(x = v, z = c_K)$$

- •特定の変数がとりうるすべての値についてその確率を足し合わせることを、その変数を周辺化marginalizationすると言う
- 得られた確率p(x)を、周辺確率と言う

## 例) ハンバーガー

	体重が 標準体重以下	体重が 標準体重以上
ハンバーガーを 毎日食べる	2	8
ハンバーガーを 毎日は食べない	38	2

## 例)確率変数を使って書くと…

	x = a	x = b
z = s	2	8
z = t	38	2

#### 問題1-1

- p(x=a)を求めよ。
- p(z=t)を求めよ。
- p(x = a, z = s)を求めよ。
- p(x = a, z = t)を求めよ。

	x = a	x = b
z = s	2	8
z = t	38	2

## 例) それぞれの場合の確率を求めると…

	x = a	x = b
z = s	$\frac{1}{25}$	4 25
z = t	$\frac{19}{25}$	$\frac{1}{25}$

## 条件付き確率 conditional probability

$$p(x = v | z = c)$$

•上は、確率変数zが値cをとるという事実が所与のとき、確率変数xが値vをとる確率(定義は下記のとおり)

$$p(x = v|z = c) \equiv \frac{p(x = v, z = c)}{p(z = c)}$$

•  $\sum_{x} p(x|z=c) = 1$ を満たす

#### 問題1-2

- p(x = a|z = s)を求めよ。
- p(x = a|z = t)を求めよ。
- p(z = s | x = a)を求めよ。
- p(z=t|x=a)を求めよ。

	x = a	x = b
z = s	2	8
z = t	38	2

#### 周辺確率と条件付き確率の関係

$$p(z=s)$$

• 上は、xの値が未知のとき、z = sである確率

$$p(z = s | x = b)$$

• 上は、x = bが観測されているとき、z = sである確率

• この2つの確率p(z)とp(z|x)の関係は $? \rightarrow ベイズ則$ 

## ベイズ則 Bayes rule

$$p(z|x) \propto p(x|z)p(z)$$

- 出来事の観測で確率が変わるという式
  - ある仮説が成り立つ確率は…
  - その仮説が成り立っていると<u>尤もらしくなる</u>出来事が観測されると…
  - 高い値になる

## ベイズ則の証明

- 条件付き確率の定義から $p(z|x) = \frac{p(x,z)}{p(x)}$
- 条件付き確率の定義から $p(x|z) = \frac{p(x,z)}{p(z)}$
- ・組み合わせると $p(z|x) = \frac{p(x,z)}{p(x)} = \frac{p(x|z)p(z)}{p(x)} \propto p(x|z)p(z)$

## ベイズ則の「比例する∝」という記号

$$p(z|x) \propto p(x|z)p(z)$$

- ∝は「左辺が右辺に比例する」という意味
- p(z|x)をちゃんと求めるには、比例定数を求める必要がある
- $\sum_{z} p(z|x) = 1$ が満たされるように、比例定数を決める

$$p(z|x) = ???$$

#### ベイズ則を等号で書くと・・・

$$p(z|x) = \frac{p(x|z)p(z)}{\sum_{z} p(x|z)p(z)}$$

• 右辺の分母はxの周辺確率、つまりp(x)に等しい

$$\sum_{z} p(x|z)p(z) = \sum_{z} \frac{p(x,z)}{p(z)} p(z) = \sum_{z} p(x,z) = p(x)$$

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

#### 例) ハンバーガー

•p(z=s)は、z=sであるとき尤もらしくなるデータが 観測されると、高くなる

$$p(z = s | x = b) \propto p(x = b | z = s)p(z = s)$$

$$p(z = s | x = b) \propto \frac{8}{10} \times \frac{1}{5}$$

	x = a	x = b
z = s	2	8
z = t	38	2

- 「∝ | は比例するという意味
- 比例関係をもとにp(z=s|x=b)の値を求めるには

$$\sum_{z} p(z|x=b) = 1$$
を使う(確率は、すべての場合について足すと1)

$$p(z = s | x = b) \propto p(x = b | z = s)p(z = s) \downarrow \emptyset$$

$$p(z = s | x = b) \propto \frac{8}{10} \times \frac{1}{5} = \frac{4}{25}$$

$$p(z = t | x = b) \propto p(x = b | z = t)p(z = t) \downarrow \emptyset$$

$$p(z = t | x = b) \propto \frac{2}{40} \times \frac{4}{5} = \frac{1}{25}$$

ということはp(z = s | x = b)の実際の値は

	4		
n(z-c x-h)	<u>25</u>	_	4
p(z=s x=b) =	$\frac{4}{4}$	_	5
	$25 \ \ 25$		

	x = a	x = b
z = s	2	8
z = t	38	2

## 確率分布

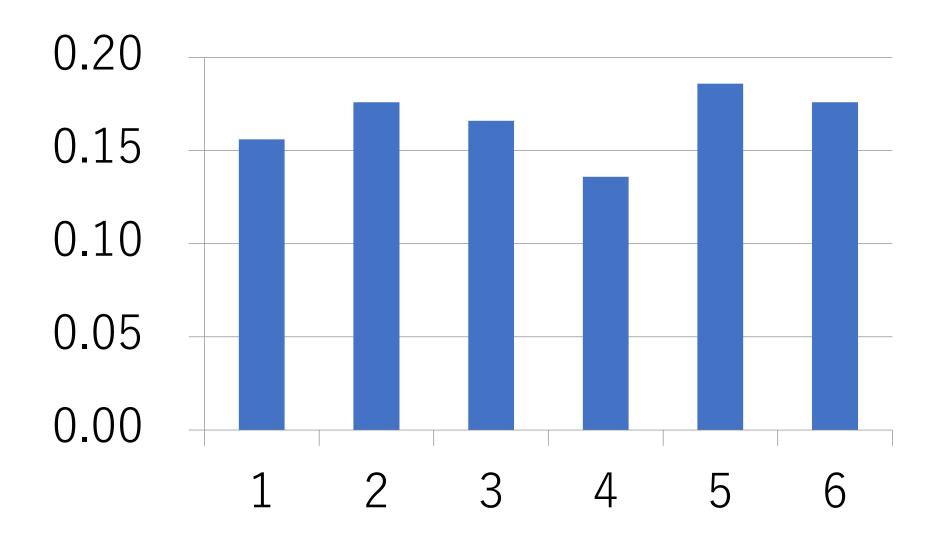
- ・確率変数について、それがとりうる値のすべてについて、その値をとる確率を示したものを、確率分布と呼ぶ
- ・確率変数がとりうる値すべてについて、その値をとる確率を足 し合わせると、1になる
- 連続値をとる確率変数の場合は、和ではなく積分で考える

離散確率分布 discrete probability distribution

#### 例) サイコロ

- 事象:「1の目が出る」, 「2の目が出る」等
- 確率分布:各事象に確率を定めたもの
  - 離散的なので事象を一個一個と数えられる
  - すべての事象の確率を足すと1になる

## 例)サイコロの目の確率分布



# 連続確率分布 continuous probability distribution

#### 例) 体重測定

- 事象:「体重が57.5kg」, 「体重が70.1kg」等
- •確率分布:<u>事象の集合</u>に確率を定めたもの
  - 連続的なので、事象を一個一個と個別に扱えない
  - すべての事象を含む集合の確率は1

#### 正規分布

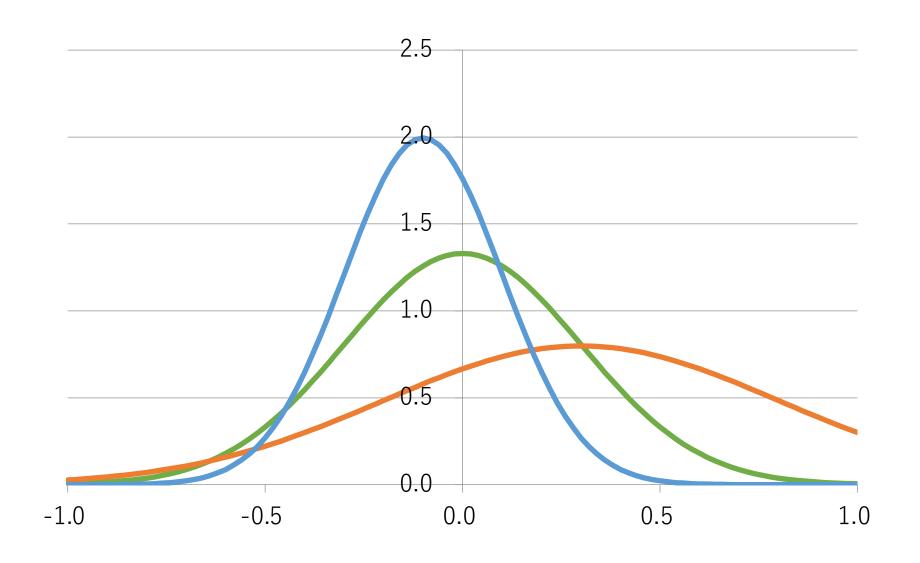
- 平均 $\mu$ と標準偏差 $\sigma$ で決まる分布
- 正規分布をあらわす関数

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

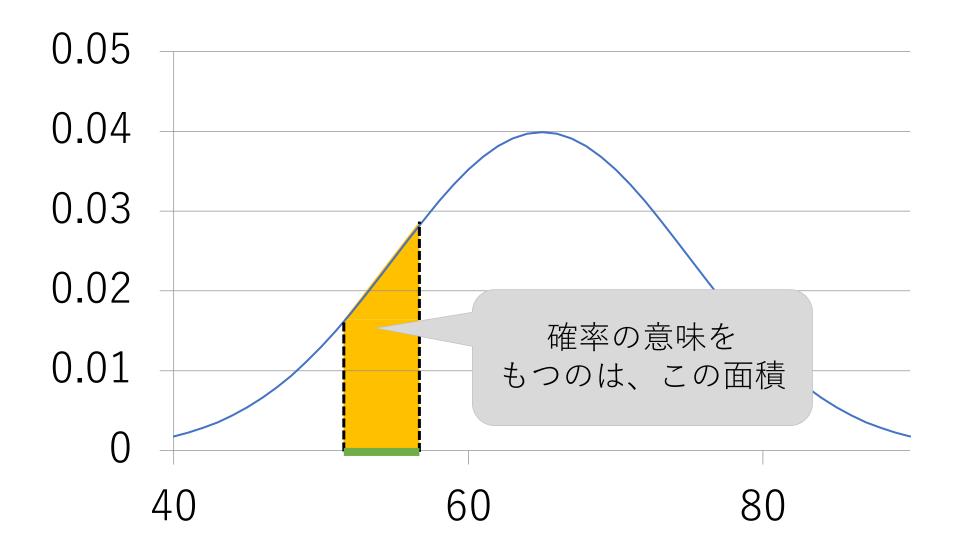
• これをある範囲で積分するとその範囲の値をとる確率が求まる

例) $\int_{45}^{55} f(x) dx$ は「体重が $45\sim55$ kg」という事象の確率

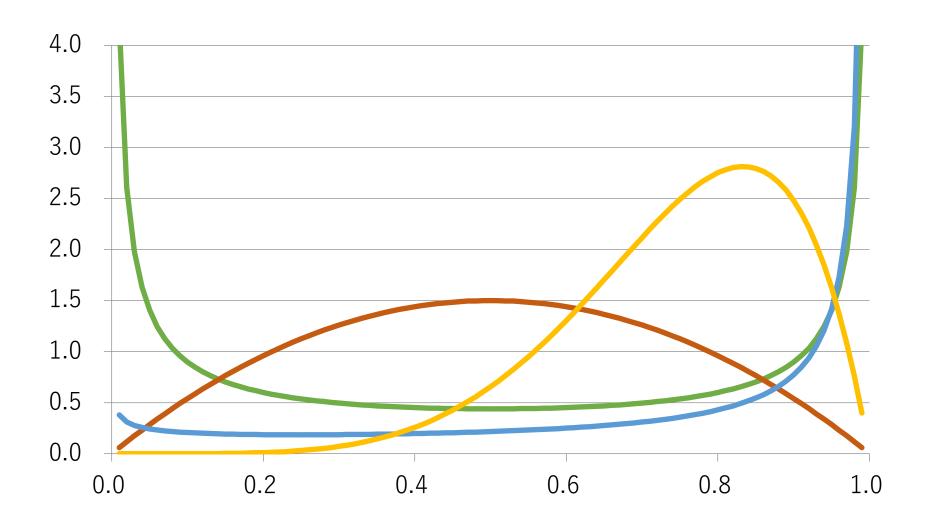
## 正規分布 normal distribution



## 例) 体重の確率分布



# ベータ分布 Beta distribution



#### 期待值

- ・確率変数xの期待値とは、<u>その変数がとりうる値とその値をと</u>る確率との積を、とりうる値すべてにわたって加算したもの
- 連続値をとる確率変数の場合は、加算するのではなく積分する 離散確率変数の場合は $\sum_x xp(x)$ 、連続確率変数の場合は $\int xp(x)dx$
- 変数xの関数f(x)についても、各々の値をとる確率を掛けて和 や積分をとることで期待値が得られる(恒等関数の場合が上のケース) 離散確率変数の場合は $\sum_x f(x)p(x)$ 、連続確率変数の場合は $\int f(x)p(x)dx$

#### 問題1-3

- ・4枚のフェアなコインを投げたとき、表が出た枚数と裏が出た枚数の差の絶対値の期待値はいくら?
  - フェアなコインとは、表が出る確率も、裏が出る確率も、 ぴったり0.5のコインのことをいう

## 問題1-3の答え

$$\sum_{i=0}^{1} |i - (4 - i)| \times \frac{4!}{i! (4 - i)!} \times \left(\frac{1}{2}\right)^4$$

$$= \left(\frac{1}{2}\right)^4 \{4 \times 1 + 2 \times 4 + 0 \times 6 + 2 \times 4 + 4 \times 1\} = \frac{24}{16} = \frac{3}{2}$$

#### 課題1-2

- バスにわんこが20匹、にゃんこが15匹乗っている。
- わんこ20匹中、12匹が白い毛である。
- にゃんこ15匹中、4匹が白い毛である。

- このバスの中から無作為に1匹、乗客を選ぶ。
  - 1. 選んだ乗客が白い毛である確率を求めよ。
- 2. 選んだ乗客がわんこであったとき、その乗客が白い毛である確率を求めよ。

