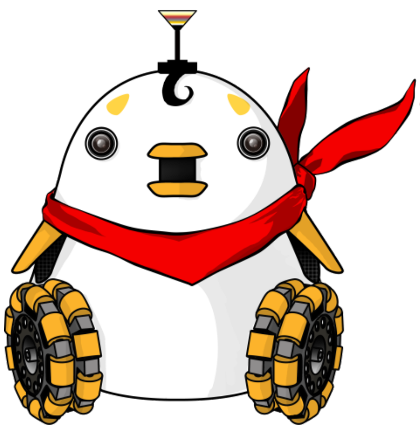
人工知能
第５回　確率モデル（1）：
確率とベイズ理論の基礎

立命館大学 情報理工学部
谷口彰



STORY 確率とベイズ理論の基礎

- これまでホイールダック２号は,
  - 自分が「左に行こう」と望めば必ず左に行けるし，「前に進もう」と望めば前に進めると思っていた.
  - 宝箱を発見するときも，宝箱の見た目は常に同じで，宝箱があれば,「あ，宝箱だ！」と確実に認識できるものだと思っていた.
- しかし，現実はそうではなかった.
  - ホイールダック２号が前進したつもりでも，オムニホイールがスリップして前に進めなかったり，左に移動しようとしても，地面のゴミを踏んでしまい車輪の一つが空転し方向がずれてしまったりした.
  - 宝箱の画像も光の当たり方や宝箱の向きなどによって毎回異なっていた. ただ宝箱の画像を持ち，その画像とピッタリ一致するものを宝箱と思えばいいと考えるのは**大きな誤り**だった.
- 世の中は**不確実性**に満ちていた．現実は秩序立った確定システムではない．**未来は確率的にしか予測できず，間違いの可能性に満ちた確率システム**だった.

仮定 確率とベイズ理論の基礎

- ホイールダック2号は過去の経験から確率の計算ができるものとする.

Contents

6.1.1 実世界の不確実性と確率

- 実世界の不確実性
  - 実世界とコンピュータ・シミュレーションの世界の違い
  - 例）
    - ボールの放物運動
    - 電子メールにおけるスパムメール
    - カメラに写る犬の画像

トーマス・ベイズ（Thomas Bayes）
1702年 - 1761年
イギリスの長老派の牧師・数学者、哲学者

- ベイズ理論(Bayes’ theory)
  - ベイズの定理を活用しながら確率論の枠組みに基づき，データからの推定や決定，解析を行う広範な理論枠組みのこと．

Contents

6.2.1 ホイールダック 2 号の不確実な前進

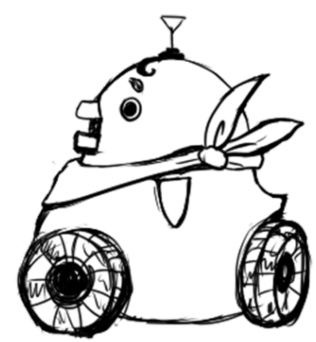
- 事象(event)
- 確率(probability)
  - 全ての事象について足し合わせると 1 になる.

$$\sum_A P(A) = 1, \quad 0 \leq P(A) \leq 1 \qquad (6.1)$$

$$\left((20+8)+(12+60)\right)/(20+8+12+60) = 1$$

表 6.1 ホイールダック 2 号の前進の履歴

出した命令 \ 実現した結果	結果：前進	結果：停止
命令：前進	20 回	8 回
命令：停止	12 回	60 回



6.2.3 同時確率(joint probability)

- 事象*A* と事象*B* がともに起こる確率
- P(“命令：前進”, “結果：前進”)

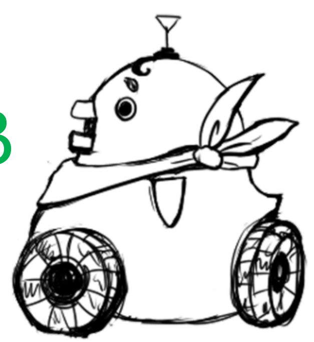
$= 20/(20+8+12+60) = 20/100 = 1/5$

表 6.1 ホイールダック 2 号の前進の履歴

出した命令 \ 実現した結果	結果：前進	結果：停止
命令：前進	20 回	8 回
命令：停止	12 回	60 回

B

A



6.2.4 加法定理(addition theorem)：和事象の確率

■加法定理

和事象（sum event）

積事象（product event）

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (6.2)$$

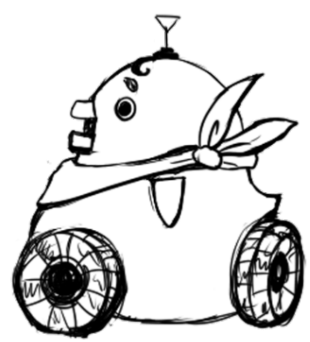
表 6.1 ホイールダック 2 号の前進の履歴

出した命令 \ 実現した結果	結果：前進	結果：停止
命令：前進	20 回	8 回
命令：停止	12 回	60 回

B

A

P(“命令：前進”, “結果：前進”)
= 28/100 + 32/100 − 20/100



6.2.5 条件付き確率 (conditional probability)

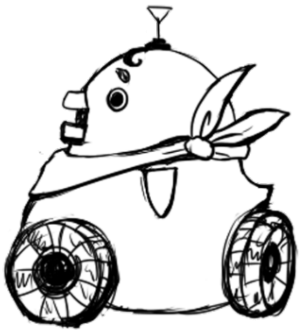
- 事象A が事象B が起こった条件のもとに起こる確率
- P(A|B): 条件付き確率(conditional probability) or 事後確率(posterior probability)
- P(“結果：前進”|“命令：前進”)
  = 20/(20+8)

表 6.1 ホイールダック 2 号の前進の履歴

出した命令 \ 実現した結果	結果：前進	結果：停止
命令：前進	20 回	8 回
命令：停止	12 回	60 回

B

A



6.2.5 乗法定理(multiplication theorem)

- 乗法定理

$$P(A,B) = P(A|B)P(B) \tag{6.2}$$

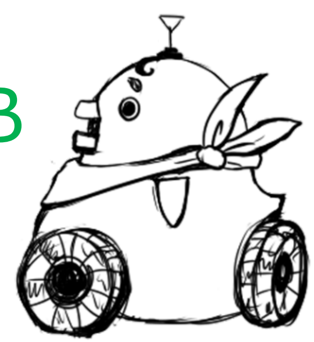
$$\frac{20}{100} = \frac{20}{\cancel{28}} \times \frac{\cancel{28}}{100}$$

表 6.1　ホイールダック 2 号の前進の履歴

出した命令 \ 実現した結果	結果：前進	結果：停止
命令：前進	20 回	8 回
命令：停止	12 回	60 回

B

A



6.2.7 周辺化(marginalization)

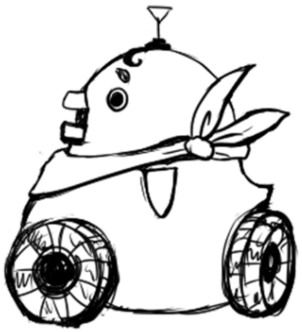
- 周辺化

$$P(A) = \sum_{B} P(A, B) \qquad (6.7)$$

表 6.1　ホイールダック 2 号の前進の履歴

出した命令 \ 実現した結果	結果：前進	結果：停止	
命令：前進	20 回	8 回	A
命令：停止	12 回	60 回	
	B	B	

$$\frac{28}{100} = \frac{20}{100} + \frac{8}{100}$$



※これを加法定理と呼ぶ教科書もある。

確率の基本式

確率の基本式

事象 A の確率	$P(A)$	(6.11)
同時確率	$P(A, B)$	(6.12)
事象 A の条件付き確率	$P(A\|B)$	(6.13)
乗法定理	$P(A, B) = P(A\|B)P(B)$	(6.14)
加法定理	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	(6.15)
周辺化	$P(A) = \sum_{B} P(A, B)$	(6.16)

演習6-1

- ２つの袋があり，皮の袋が2/3の確率で選ばれる．布の袋が1/3の確率で選ばれる．それぞれの袋には下記の玉がそれぞれ入っており，袋が選ばれるとその後は全ての玉が等確率で取り出される．以下を求めよ．

	Y1 : 赤い玉	Y2 : 青い玉	Y3 : 黄色い玉
X1 : 皮の袋	15個	5個	0個
X2 : 布の袋	15個	1個	4個

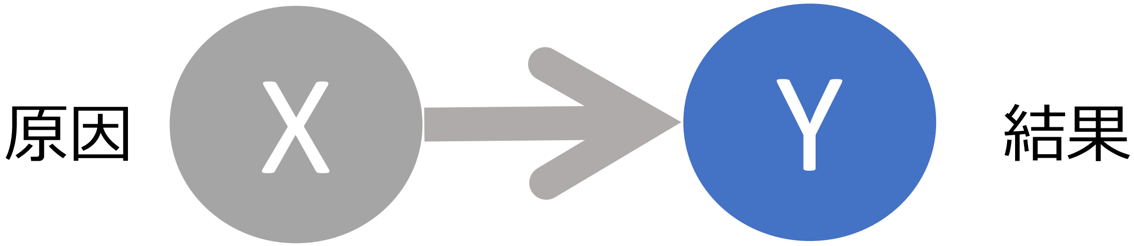
1. P(X1)
2. P(Y2|X2)
3. P(X1,Y2)
4. P(Y1)

Contents

- 6.1 環境の不確実性
- 6.2 確率の基礎
- 6.3 ベイズの定理
- 6.4 期待値と意思決定
- 6.5 確率分布のパラメータ推定

確率の取り扱いとベイズ理論

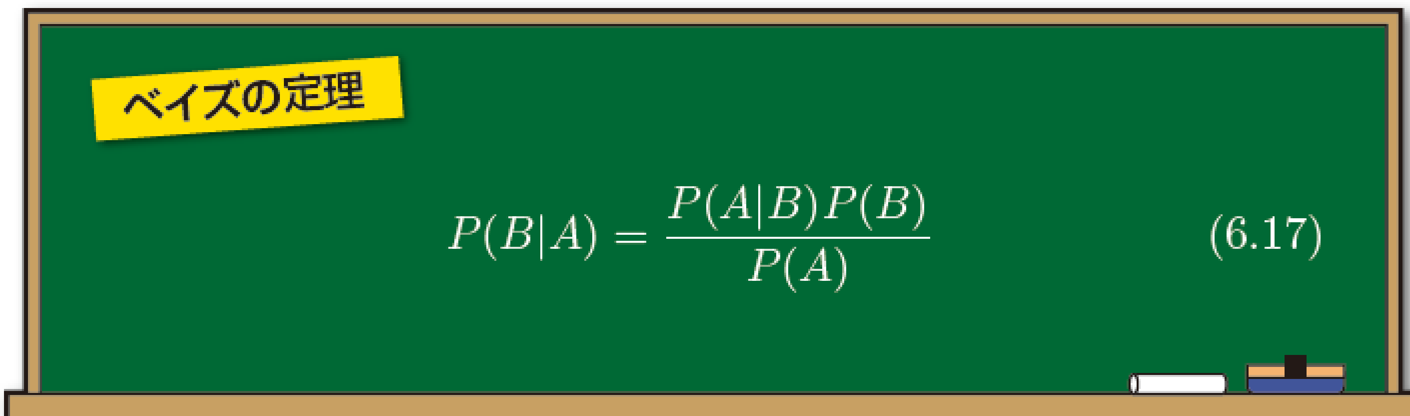
- ある事象が起こり，その原因としていくつかの事象が考えられ，それらは互いに独立な事象であり，それぞれがある確率をもって起こるとする.
- このときベイズ理論では結果として起こった事象に対する原因がどれであったかという確率を求める事ができる.



非常に柔軟な枠組みであり，機械学習，自然言語処理，パターン認識，音声認識はじめ，多くのデータを扱う情報処理で一般的に用いられるようになっている.

6.3.1 ベイズの定理の導出

- 実際のところは条件付き確率の性質から自然と導かれる基本的な式



$$P(B|A)P(A) = P(A,B) = P(A|B)P(B) \tag{6.18}$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{\sum_{B'} P(A,B')} = \frac{P(A|B)P(B)}{\sum_{B'} P(A|B')P(B')} \tag{6.19}$$

実用上、分母が直接求まらないことが多いので、(6.19)のように分解することが多い

隠れた事象の推定とベイズの定理

- 例えば，$C = \{C_1, C_2, \ldots, CK\}$ のいずれかの事象が生じる場合を考える．

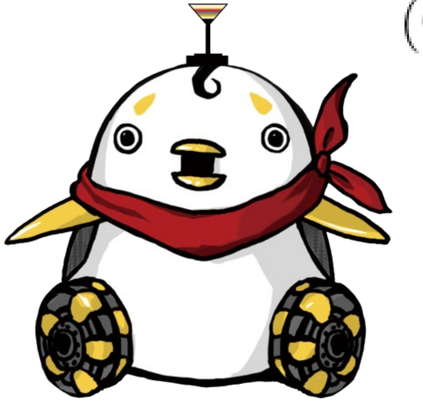
$$\overbrace{P(C_j|A)}^{\text{事後確率}} = \frac{P(A, C_j)}{P(A)} = \frac{P(A, C_j)}{\sum_k P(A, C_k)} \tag{6.19}$$

$$= \frac{\overbrace{P(A|C_j)}^{\text{尤度}}\overbrace{P(C_j)}^{\text{事前確率}}}{\sum_{C_k} P(A|C_k)P(C_k)} \tag{6.20}$$

$$\propto P(A|C_j)P(C_j) \tag{6.21}$$

C_1: 罠

C_2: 宝



6.3.2 ベイズの定理の意味

- **原因と結果の関係を逆転**させることができるのがベイズの定理の主要な機能である.
  - 例）濡れている地面（結果：観測）を見て雨が降ったか（原因：隠れた事象）どうかを考える
  - 例）ホイールダック2 号が前進したの（結果）を見て「果たしてホイールダック2 号は前進命令を出した（原因）のか？」と考える

表 6.2　確定システムと不確実を有するシステムにおける原因結果の捉え方の違い

確定システム	不確実性を有するシステム
$B \to A$	$P(A\|B)$

6.3.3 ホイールダック2号の動作の原因を推定する

- ホイールダック2 号が前進しているのが観測された時（A=”結果：前進”）に，実際にホイールダック2 号が前進命令（B=”命令：前進”）を出している確率はいくらか？

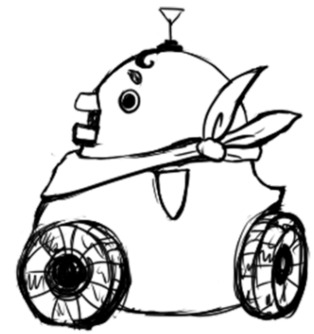


表 6.3　ホイールダック 2 号の行動選択とその結果の確率

出した命令 B	事前確率 $P(B)$
命令:前進	0.3
命令:停止	0.7

実現した結果 A ＼ 出した命令 B	命令 B が与えられたときに結果 A が観測される確率（尤度）： $P(A\|B)$	
	結果:前進	結果:停止
命令:前進	0.8	0.2
命令:停止	0.4	0.6

$$
\begin{aligned}
&P(B=\text{“命令:前進”}|A=\text{“結果:前進”})\\
&=\frac{P(A=\text{“結果:前進”}|B=\text{“命令:前進”})P(B=\text{“命令:前進”})}{\sum_{B'}P(A=\text{“結果:前進”}|B')P(B')}\\
&=\frac{P(A=\text{“結果:前進”}|B=\text{“命令:前進”})P(B=\text{“命令:前進”})}{\left\{\begin{array}{c}P(A=\text{“結果:前進”}|B'=\text{“命令:前進”})P(B'=\text{“命令:前進”})\\ +P(A=\text{“結果:前進”}|B'=\text{“命令:停止”})\end{array}\right\}}\\
&=\frac{0.8\times 0.3}{0.8\times 0.3+0.4*0.7}=24/52\simeq 0.46 \qquad (6.20)
\end{aligned}
$$

$$
\begin{aligned}
&P(B=\text{“命令:停止”}|A=\text{“結果:前進”})\\
&=\frac{P(A=\text{“結果:前進”}|B=\text{“命令:停止”})P(B=\text{“命令:停止”})}{\sum_{B'}P(A=\text{“結果:前進”}|B')P(B')}\\
&=\frac{P(A=\text{“結果:前進”}|B=\text{“命令:停止”})P(B=\text{“命令:停止”})}{\left\{\begin{array}{c}P(A=\text{“結果:前進”}|B'=\text{“命令:前進”})P(B'=\text{“命令:前進”})\\ +P(A=\text{“結果:前進”}|B'=\text{“命令:停止”})\end{array}\right\}}\\
&=\frac{0.4\times 0.7}{0.8\times 0.3+0.4*0.7}=28/52\simeq 0.54 \qquad (6.21)
\end{aligned}
$$

※分母が同じになることに注目！
※「分子」の比さえわかればよい！

演習 6-2

- ２つの袋があり，皮の袋が2/3の確率で選ばれる．
  布の袋が1/3の確率で選ばれる．それぞれの袋には下記の玉がそれぞれ入っており，袋が選ばれると，その後は全ての玉が等確率で取り出される．

	Y1 : 赤い玉	Y2 : 青い玉	Y3 : 黄色い玉
X1 : 皮の袋	15個	5個	0個
X2 : 布の袋	15個	1個	4個

1. P(X1|Y2)を求めよ．および P(X2|Y2)を求めよ．
2. P(X2|Y3)を求めよ．
3. 赤い玉が取り出された時（結果），取り出した袋はどちらだった（原因）可能性が高いか？

Contents

6.4.1 確率変数と期待値

□確率変数xと関数fの期待値

$$E[x] = \sum_x xP(x) \tag{6.25}$$

$$E[f] = \sum_A f(A)P(A) \tag{6.26}$$

□関数fの条件付き期待値

$$E[f|B] = \sum_A f(A)P(A|B) \tag{6.28}$$

- 簡単な意思決定問題を考える上で重要.
- 確定システムではこのようなことを考えなくても，sとa が定まれば次状態も利得も1 通りに決まっていたので，期待値を考える必要はなかった．しかし，不確実性を持つ実世界ではこのような確率を考えることが重要となる.

演習 6-3 期待値

- ２つの袋があり，皮の袋が2/3の確率で選ばれる．布の袋が1/3の確率で選ばれる．それぞれの袋には下記の玉がそれぞれ入っており，袋が選ばれるとその後は全ての玉が等確率で取り出される．

	Y1 : 赤い玉	Y2 : 青い玉	Y3 : 黄色い玉
X1 : 皮の袋	15個	5個	0個
X2 : 布の袋	15個	1個	4個

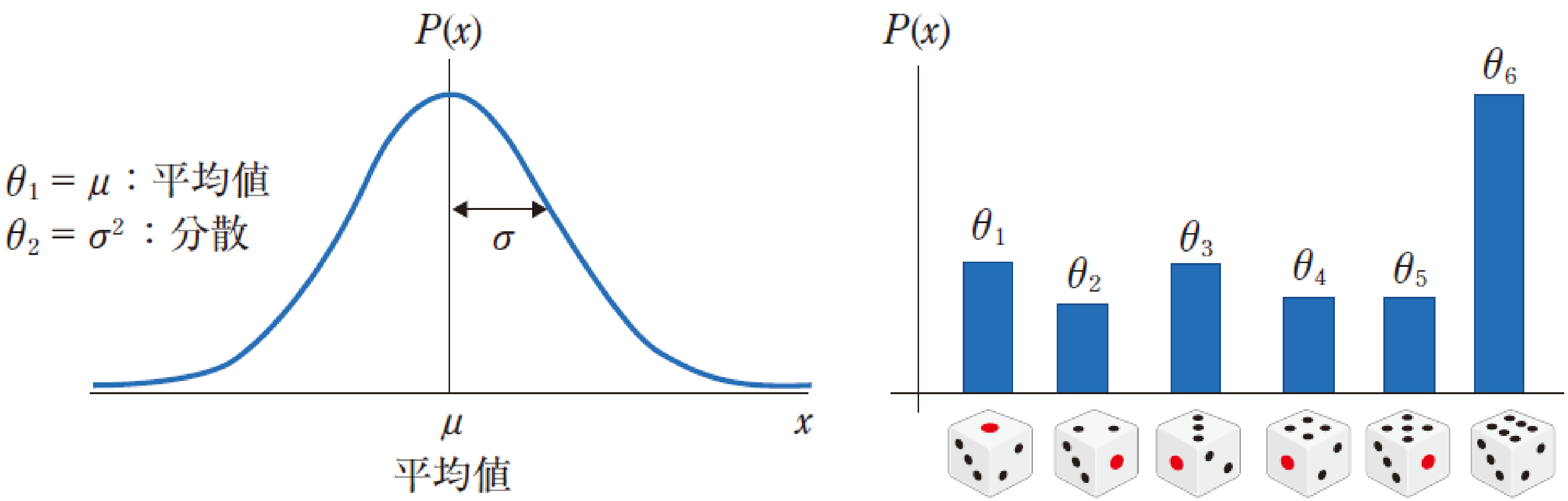
- 取り出した玉が赤い玉なら1 点，青い玉なら2 点，黄色い玉なら3 点得られる．

1. 得られる得点の期待値を求めよ．
2. 皮の袋から取り出したということがわかっている場合，玉を一つ取り出した場合に得られる得点の条件付き期待値を求めよ．

Contents

6.5.1 確率分布の推定

- 確率分布はなんらかの**パラメータ**に従う。
- これを推定することを確率分布のパラメータ推定という。



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

$$P(x) = \theta_x$$

図 6.2 確率分布とパラメータ

MLE: maximum likelihood estimation/estimator

6.5.2 最尤推定

- 確率分布のパラメータ推定に用いられる方法の一つが最尤推定(maximum likelihood estimation) であり，得られた観測に対して最も尤もらしいパラメータを求める．得られたデータに対して尤もらしいパラメータθ^{MLE}を求める。

$$\theta^{MLE} = \underset{\theta}{\operatorname{argmax}}\, P(X|\theta) \qquad (6.31)$$

例) ホイールダック2 号が命令の結果として前進する確率をμ とした時，合計N 回の試行を行い前進した回数がm 回だった確率は以下の二項分布で書くことができる.

$$P(m|\mu) = \frac{N!}{(N-m)!m!}\mu^m(1-\mu)^{(N-m)}$$

尤度 $L(\mu) = P(m|\mu)$ の対数 $l(\mu) = \log P(m|\mu)$ を取る.

$$l(\mu) = \log \frac{N!}{(N-m)!m!}\mu^m(1-\mu)^{(N-m)}$$

$l(\mu)' = \frac{m}{\mu} - \frac{N-m}{(1-\mu)} = 0$ が極値

$$= m\log\mu + (N-m)\log(1-\mu) + \log\frac{N!}{(N-m)!m!}$$

よって $\mu^{MLE} = \frac{m}{N}$

6.5.3 ベイズ推定

- パラメータのベイズ推定はパラメータも他の変数と同じように確率変数として捉えて，ベイズ推定の枠組みによって求めようという考え方

事前分布 $P(\theta)$

観測 $X = \{x_1, x_2, \ldots, x_N\}$

$$P(\theta|X) \propto P(X|\theta)P(\theta)$$

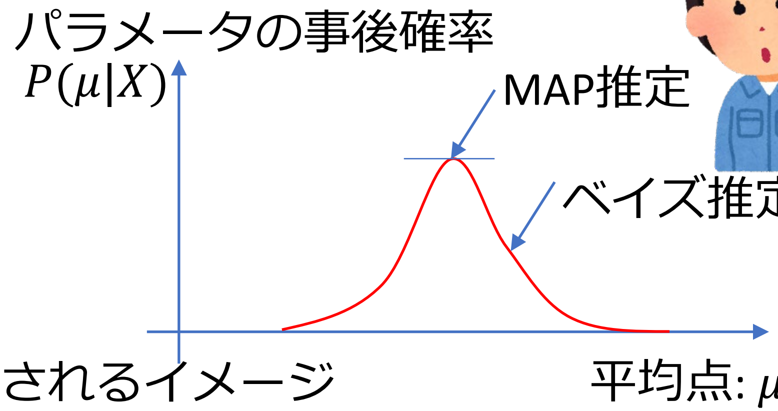
※ベイズ推定ではパラメータϑ は一点に決まるのではなく，確率分布として求まる．

クラスの平均点は50点かもしれないし51点かもしれない……

{40点, 50点, 60点}~ $P(x)$

MAP推定

$$\theta^{MAP} = \underset{MAP}{\operatorname{argmax}} P(X|\theta)P(\theta)$$



※最尤推定に事前分布が加わって補正されるイメージ

第５回のまとめ

- 環境の不確実性を取り扱うために確率を用いることの重要性を学んだ.
- 確率の基礎に関して学んだ．またベイズの定理を導入し，その意味について学んだ.
- 期待値や条件付き期待値の計算方法について学んだ.
- 最尤推定やベイズ推定による確率分布のパラメータ推定に関して学んだ.