第５章 教師付き学習と教師なし学習

國井慎也

コイン投げ問題

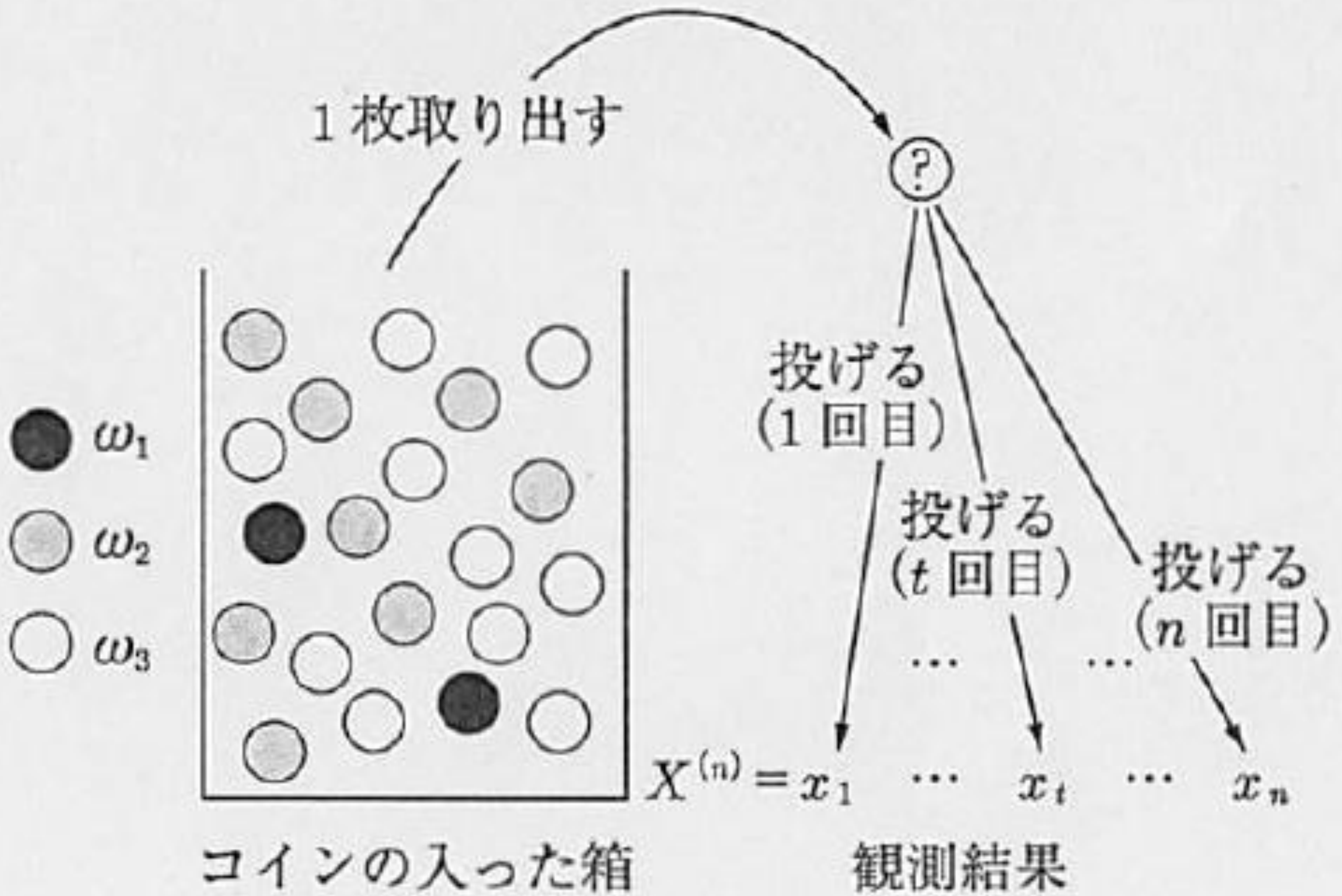


図 2·1　箱から取り出したコインを投げる実験

サイコロ投げ問題

- サイコロの問題
  - サイコロの種類　$\omega_1, \ldots, \omega_c$
  - 含有率　$\pi_1, \ldots, \pi_c$
  - サイコロω_iを投げた時kの目v_kが出る確率　$\theta_{ik} = P(v_k|\omega_i)$

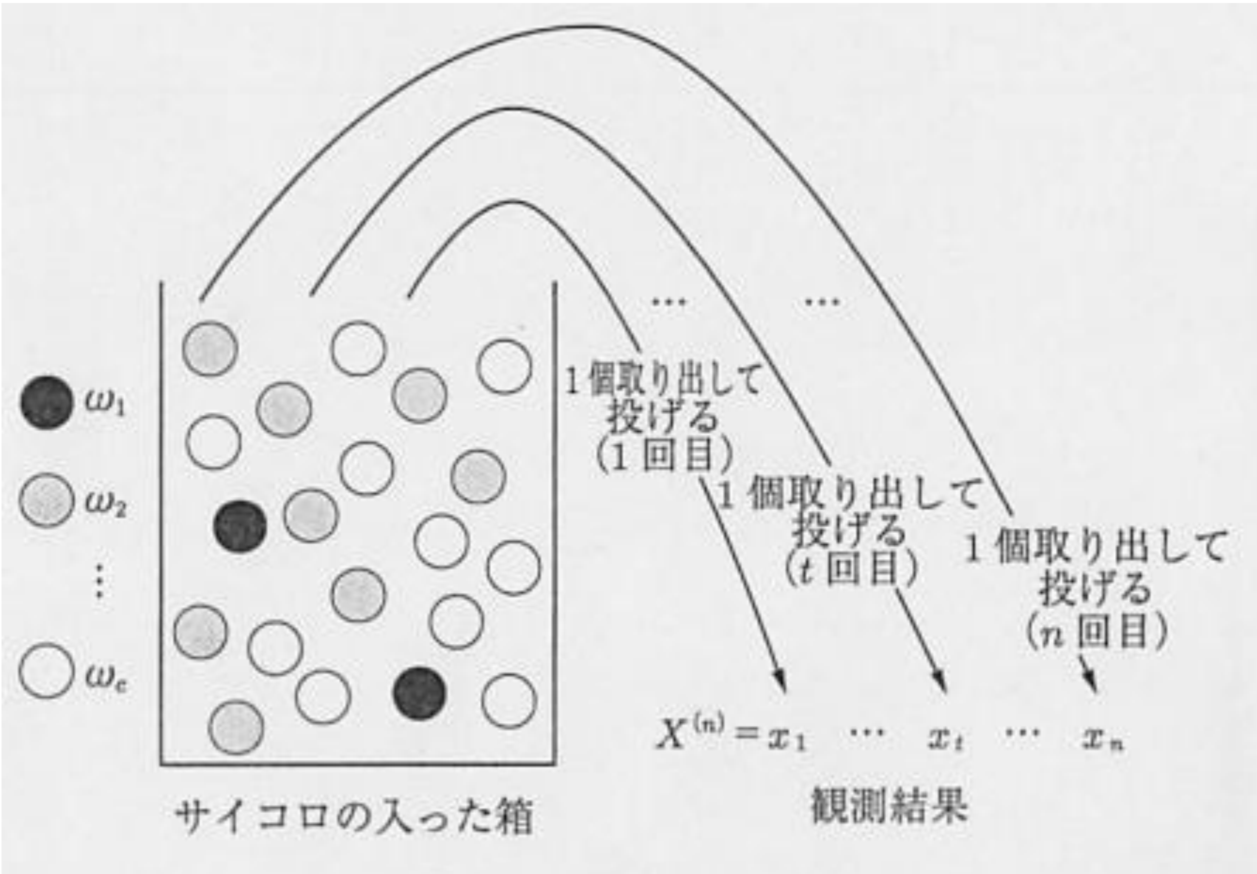
  - サイコロω_iをn回投げて観測結果$X^{(n)}$が得られる確率

$$P\left(X^{(n)}\middle|\omega_i\right) = \prod_{k=1}^{m} \theta_{ik}^{rk}$$

  - ベイズ識別関数

$$g_i\left(X^{(n)}\right) = P(\omega_i) * P\left(X^{(n)}\middle|\omega_i\right) = \pi_i \prod_{k=1}^{m} \theta_{ik}^{rk}$$

サイコロ投げ問題



ω_1
ω_2
ω_c
1個取り出して
投げる
(1回目)
1個取り出して
投げる
(t回目)
1個取り出して
投げる
(n回目)
$X^{(n)} = x_1 \cdots x_t \cdots x_n$
サイコロの入った箱
観測結果

サイコロ投げ問題

- 観測結果 X

$$X = x_1 x_2 \dots x_t \dots x_n \qquad x_t \in \{\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_\mathrm{k}\}$$

- 結果に対するサイコロの種類S

$$\mathrm{S} = s_1 s_2 \dots s_t \dots s_n \qquad s_t \in \{\omega_1, \omega_2, \cdots, \omega_\mathrm{c}\}$$

- サイコロの種類系列Sがある場合
  - 教師付き学習
  - 系列S　教師信号
  - 結果xと系列Sを完全データ
- 系列Sがない場合
  - 教師なし学習
  - 結果xのみを不完全データ

最尤推定のための数学

- 定理5.1
  - n個の定数ω$_1$, ω$_2$, ⋯ , ω$_n$がある。ここで、n個の変数 $x_1 x_2 \ldots x_n (0 \leq x_n \leq 1)$が、拘束条件
$$\sum_{i=1}^{n} x_i = 1$$
    を満たす。このとき、
$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \omega_i log\, x_n$$
    を最大にするx_iは
$$x_i = \frac{\omega_i}{\sum_{k=1}^{n} \omega_k}$$

教師あり学習

- サイコロの問題
  - サイコロの種類　$\omega_1, \dots, \omega_c$
  - 含有率　$\pi_1, \dots, \pi_c$
  - サイコロω_iを投げた時kの目v_kが出る確率　$\theta_{ik} = P(v_k|\omega_i)$
  - 系列$X = x_1 x_2 \dots x_t \dots x_n$
  - $\mathrm{S} = s_1 s_2 \dots s_t \dots s_n$
- パラメータ推定
  - $\hat{\theta} = argmax_\theta \{logP(X,S)\}$
  - $logP(X,S) = logP(S) + \log(\mathrm{X}|\mathrm{S})$
  - $= \sum_{t=1}^{n} \log P(s_t) + \sum_{t=1}^{n} \log P(x_t|s_t)$

パラメータπ_iの推定

- 第一項

$P(s_t)$は$P(w_1),..,\mathrm{P(w_c)}$つまり$\pi_1,..,\pi_c$

$$\sum_{t=1}^{n} \log P(s_t) = \sum_{k=i}^{c} n_i \log \pi_i$$

n_i:サイコロw_iを取り出した回数

定理5.1より、π_iの推定値は

$$\hat{\pi}_i = \frac{n_i}{\sum_{j=1}^{c} n_j} = \frac{n_i}{n}$$

相対頻度を求めている

パラメータθ_{ik}の推定

- 第二項目

$$\sum_{t=1}^{n} \log P(x_t | s_t) = \sum_{i=1}^{c} \sum_{k=1}^{m} n_{ik} \log \theta_{ik}$$

θ_{ik}:サイコロw_iの観測結果がv_kである確率
n_{ik}:サイコロw_iの観測結果がv_kだった回数

この式を最大化するには、

$$\sum_{k=1}^{m} n_{ik} \log \theta_{ik}$$

定理5.1より

$$\hat{\theta}_{ik} = \frac{n_{ik}}{\sum_{j=1}^{n} n_{ij}} = \frac{n_{ik}}{n_i}$$

教師なし学習

- サイコロの問題
  - サイコロの種類 $\omega_1, \dots, \omega_c$
  - 含有率 $\pi_1, \dots, \pi_c$
  - 観測結果系列X=$x_1 x_2 \dots x_t \dots x_n$
- 全体の観測結果の確率
  - $P(X) = P(x_1, x_2, \dots, x_t, \dots, x_n) = \prod_{t=1}^{n} P(x_t)$
  - $logP(X) = \sum_{t=1}^{n} logP(x_t)$
  - $= \sum_{k=1}^{m} r_k \, logP(v_k)$

r_k:観測結果Xのうちv_kが出た回数

- 定理5.1より
  - $P(v_k) = \frac{r_k}{\sum_{l=1}^{m} r_1} = \frac{r_k}{n} = \sum_{i=1}^{c} \pi_i \, \theta_{ik}$

パラメータπ_iの推定

- ラグランジュ未定乗数法
  - $L \equiv logP(X) - \lambda(\sum_{i=1}^{c} \pi_i - 1)$
  - $\frac{\partial L}{\partial \pi_i} = 0$
- 推定値
  - $\hat{\pi}_i = \frac{1}{n}\sum_{t=1}^{n} P(\omega_i|x_t) = \frac{1}{n}\sum_{k=1}^{m} r_k * P(\omega_i|v_t)$
  - $= \frac{\text{投げたサイコロが}\omega_i\text{であった回数の期待値}}{\text{サイコロを投げた回数}}$

パラメータθ_{ik}の推定

- ラグランジュ未定乗数法
  - $L \equiv logP(X) - \lambda(\sum_{k=1}^{m} \sum_{i=1}^{c} \pi_i * \theta_{ik} - 1)$
  - $\frac{\partial L}{\partial \theta_i} = 0$

- 推定値
  - $\hat{\theta}_{ik} = \frac{r_k * P(\omega_i | v_k)}{\sum_{l=1}^{m} r_l * P(\omega_i | v_l)} = \frac{\text{観測結果が} v_k \text{で、投げたサイコロが} \omega_i \text{であった回数の期待値}}{\text{投げたサイコロが} \omega_i \text{であった回数の期待値}}$

実験

- サイコロω1, ω2, ω3の含有率がそれぞれ π_1, π_2, π_3、偶数になる確率が0.8 , 0.6 , 0.3とする。サイコロを無造作に一個投げ、出た目の偶奇を観測する。
  - 試行回数10000回(r_1(奇数) $= 4746, r_2$(偶数) $= 5254$)
  - θ_{ik}は既知、π_iを推定
  - π_i の推定にEMアルゴリズムを利用
  - 初期値$(\pi_1, \pi_2, \pi_3) = (0.3,0.5,0.2)$

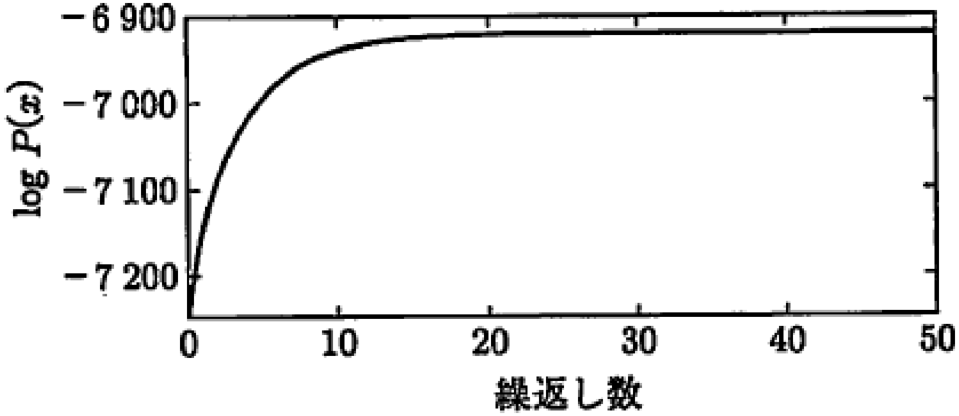


図 5·2　単調増加する $\log P(\mathrm{x})$

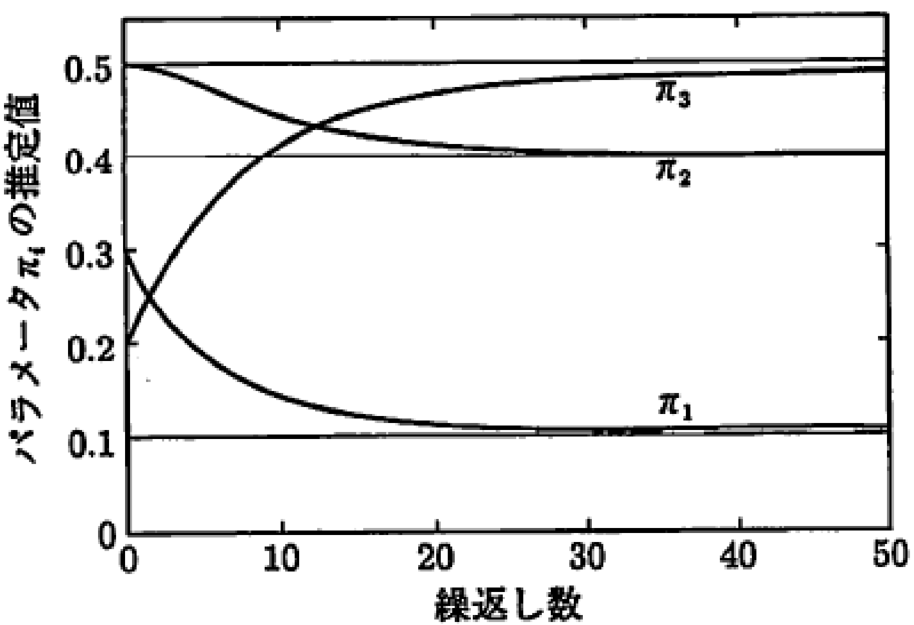


図 5·3　パラメータ π_i の推定実験