入門ベイズ統計

第4章　ベイズ判別問題とパターン認識

10月**3**日（金）
豊川　幸秀

流れ

- 4.1 メッセージと符号
- 4.2 ベイズ決定
- 4.3 正規分布をもつノイズ

メッセージと符号

- 通信理論について

  1.ベイズ決定の理解と発展に役立つ

  2.通信システムの理解に役立つ

→1から、次節に対する導入に位置する

用語等

メッセージ	符号
「イエス」	＋＋－－－
「ノー」	＋－＋－＋

・メッセージ → 符号：
符号化 or エンコーディング

・符号 → メッセージ：
復号化

・復号化の際、ノイズによりメッセージを誤認することがある。

↓

統計的決定を行う

ベイズ決定(1)

・送信される信号

θ_1＝(1,1,-1,-1,-1)　　　θ_2 ＝(1,-1,1,-1,1)

・受信される信号

$$x = (x_1, x_2, x_3, x_4, x_5)$$

より一般的に、

送信される信号: $\theta_1, \theta_2, \cdots, \theta_k$ (k≧2)

受信される信号: $x = (x_1, x_2, \cdots, x_p)$

ベイズ決定(2)

- x について、各θ から見た確率分布

$$f(x|\theta_1), f(x|\theta_2), \cdots, f(x|\theta_k)$$

↓

各θに対する領域を割り出す

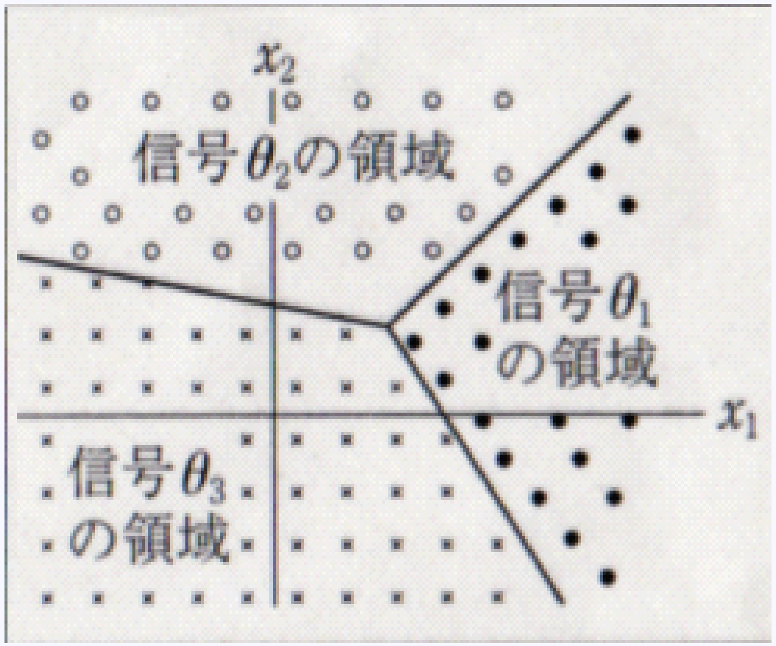

図**4.4 p=2,k=3**の場合の統計的決定の領域および決定境界

ベイズ決定(3)

- k =2,p =1とする

$$f(x|\theta_1), f(x|\theta_2)$$

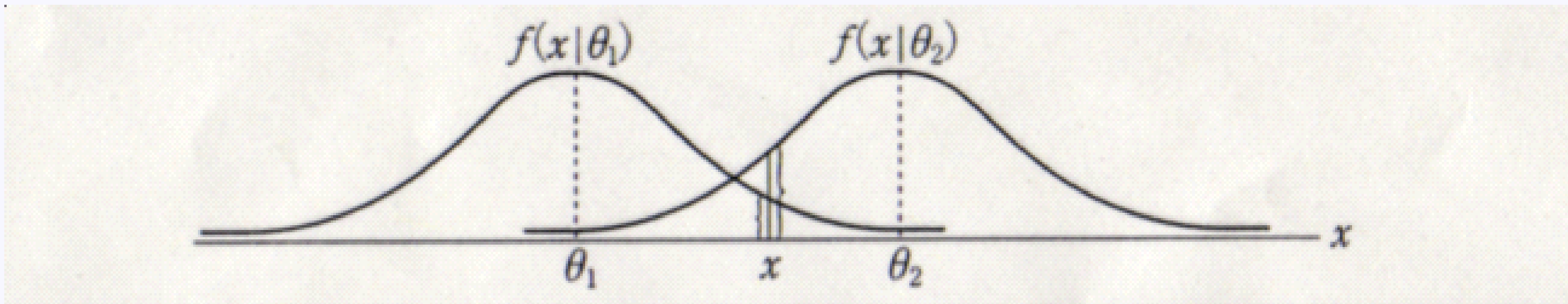

図4.5 送信される信号が**2**種類の場合

ベイズ決定(4)

直感的に、$f(x|\theta_2) > f(x|\theta_1)$ より x はθ_2と判断

↓

必ずしも正しくはない

<u>各f について、判断のコストは変わってくる</u>

送信信号＼受信者の判断	a_1（θ_1 と判断）	a_2（θ_2 と判断）
θ_1	l_{11}	l_{12}
θ_2	l_{21}	l_{22}

送信＼受信	「イエス」	「ノー」
「イエス」	0	8
「ノー」	10	0

表**4.2** 損失のマトリックス$L(\theta, a)$

ベイズ決定(5)

- θ_1, θ_2の事後確率をそれぞれw'_1, w'_2とする
  （$w'_1 > 0, w'_2 > 0, w'_1 + w'_2 = 1$）

- 2.1節の事後期待損失最小化の原理より、
  (1)a_1に対する損失の期待値

$$w'_1 l_{11} + w'_2 l_{21}$$

  (2)a_2に対する損失の期待値

$$w'_1 l_{12} + w'_2 l_{22}$$

  小さい値を与えるほうのaをとればよい。

ベイズ決定(6)

・すなわち、

$$w'_1(l_{12}-l_{11}) \le or \ge w'_2(l_{21}-l_{22})$$

・θ_1, θ_2の事前確率をそれぞれw_1, w_2とする
ベイズの定理より、$w'_i \propto w_i \cdot f(x|\theta_i)$

・よって、

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} \le or \ge \frac{w_1(l_{12}-l_{11})}{w_2(l_{21}-l_{22})} \qquad (4.2.10)$$

これが、ベイズ決定となる。

ベイズ決定(7)

(4.2.10)式の左辺を $\lambda(x)$ 、右辺を C とすると、

$\lambda(x)$: 尤度比
C : しきい値

と呼ばれる。

正規分布をもつノイズ(1)

- $f(x|\theta_i)$ は正規分布の形でよく用いられる
  (この場合、ノイズも正規分布で表現)

$$f(x|\theta_i) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\{-\frac{(x-\theta_i)^2}{2\sigma^2}\}$$

- (4.2.10)式に対数をとると、

$$\log \lambda(x) \le or \ge \log C$$

- 例として、θ_1 ＝0、θ_2 ＝1とし、正規分布N(0, σ^2)に従う分散 $\sigma^2 = \frac{1}{12}$ $(\sigma = \sqrt{\frac{1}{12}} \approx 0.290)$ のノイズが存在するとする。

正規分布をもつノイズ(2)

- 受信される信号xは以下の確率分布に従う。

$$f(x|\theta_i)=\frac{1}{\sqrt{\frac{\pi}{6}}}\cdot\exp\{-6(x-\theta_i)^2\}(i=1,2)$$

- 信号の送信頻度は、θ_1 : θ_2 ＝ 2 : 3とする。よって、事前確率は w_1＝0.400, w_2＝0.600。また、各コストは表4.2のものを使用。

これらから尤度$\lambda(x)$ を計算すると、尤度比の対数は、

$$\log\lambda(x)=6(2x-1)$$

正規分布をもつノイズ(3)

C のほうも計算していくと、最終的に、

$$x \le or \ge 0.448$$

$\theta_2 = 1$ を $\theta_1 = 0$ と誤る確率は、

$$e(0|1) = \int_{-\infty}^{0.448} \frac{1}{\sqrt{\frac{\pi}{6}}} \cdot \exp\{-6(x-1)^2\} dx \quad = 0.0278$$

逆は、 $$e(1|0) = \int_{0.448}^{\infty} \frac{1}{\sqrt{\frac{\pi}{6}}} \cdot \exp(-6x^2) dx \quad = 0.0603$$

正規分布をもつノイズ(4)

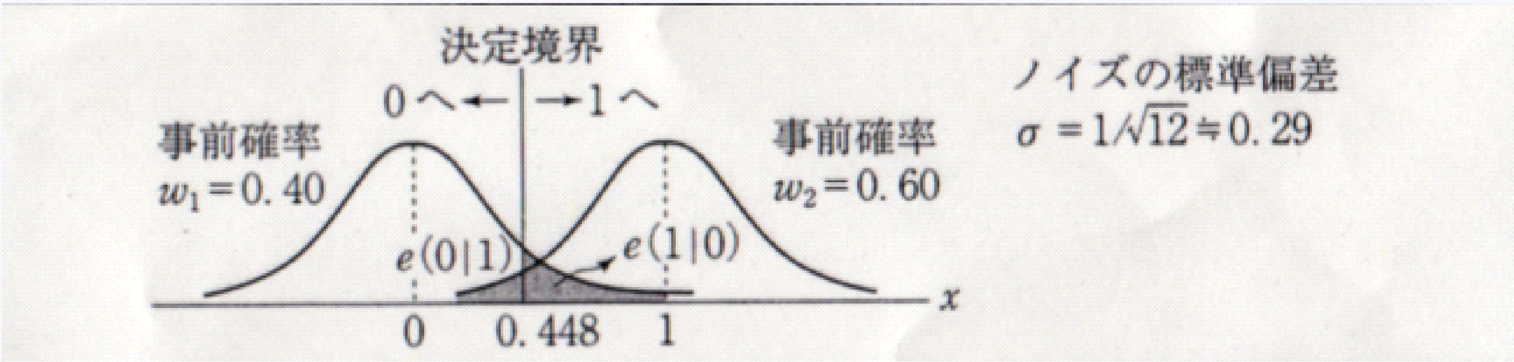

図**4.7** 送信された信号の判断

送信 ⟶ 受信	確率（%）
「イエス」 ⟶ 「イエス」	93.970
「イエス」 ⟶ 「ノー」	6.030
「ノー」 ⟶ 「ノー」	97.220
「ノー」 ⟶ 「イエス」	2.780

表**4.3** 伝送の確率１