5.　ベイズ統計

5.1 信号とシステム

ベイズの定理を導出するために必要な最小限の確率の知識

ある標本空間で定義される二つの事象A, Bの確率の定義

同時確率（joint probability） (記号 : $p(A, B)$, $p(A \cap B)$)

AとBが同時に起こる確率．結合確率とも呼ばれる．

周辺確率（marginal probability） (記号 : $p(A)$, $p(B)$)

他の事象に関わりない一つの事象だけの確率．普通の確率．

条件付確率（conditional probability） (記号 : $p(A|B)$)

Bが起こったという条件のもとでAが起こる確率．

定義された確率がもつ性質

独立（independent）

$p(A, B) = p(A)p(B)$が成り立つとき，事象Aと事象Bは独立であると言われる．独立であれば$p(A|B) = p(A), p(B|A) = p(B)$が成り立つ．

排反(exclusive)

$p(A, B) = 0$が成り立つとき，事象Aと事象Bは排反であると言われる．排反であれば，$p(A|B) = p(B|A) = 0$が成り立つ．

5.1 確率の初歩

同時確率と周辺確率の関係

	事象B_1	事象B_2		事象B_n	周辺確率
事象A_1	$p(A_1, B_1)$	$p(A_1, B_2)$	$\cdots$	$p(A_1, B_n)$	$p(A_1)$
事象A_2	$p(A_2, B_1)$	$p(A_2, B_2)$	$\cdots$	$p(A_2, B_n)$	$p(A_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
事象A_m	$p(A_m, B_1)$	$p(A_m, B_2)$	$\cdots$	$p(A_m, B_n)$	$p(A_m)$
周辺確率	$p(B_1)$	$p(B_2)$	$\cdots$	$p(B_n)$	$\sum = 1$

同時確率と周辺確率，条件付き確率の関係

$$p(A_i|B_j) = \frac{p(A_i, B_j)}{p(B_j)}$$

e.g. $$p(A_1|B_1) = \frac{p(A_1, B_1)}{p(B_1)} = \frac{1/9}{4/9} = \frac{1}{4}$$

5.2 ベイズの定理

条件付確率の定義式より

$$\begin{cases} p(A|B) = \dfrac{p(A,B)}{p(B)} & (1) \\ p(B|A) = \dfrac{p(A,B)}{p(A)} & (2) \end{cases}$$

ここで，$p(A) \neq 0$，$p(B) \neq 0$である．式(1)(2)の$p(A,B)$について以下のように変形でき，これを乗法定理という．

$$p(A,B) = p(A|B)p(B) = p(B|A)p(A) \qquad (3)$$

この関係を$p(A|B)$について変形すると

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)} \qquad (4)$$

となり，ベイズの定理が導出できる．

5.2 ベイズの定理

$$p(A|B) = p(A) \cdot \frac{p(B|A)}{p(B)}$$

$p(A)$	Aが原因で発生する事象の確率 「事前確率」もしくは「事前分布」
$\frac{p(B\|A)}{p(B)}$	新しい情報である事象Bによる修正項
$p(A\|B)$	「事後確率」もしくは「事後分布」

5.3 ベイズ統計

観測データ$\boldsymbol{y} = \{y(1), y(2), \cdots, y(n)\}$が与えられたとき，
これに対応する統計モデル（statistical model）

$$\mathcal{S} = \{p(\boldsymbol{y}, \boldsymbol{\theta})\}$$

$\mathcal{S}$: 観測データを生成したシステムの候補集合

$\boldsymbol{\theta}$: 未知パラメータ（状態量やシステムパラメータ）

$p(\boldsymbol{y}, \boldsymbol{\theta})$: 観測データの確率分布

を求める問題を考える．

ベイズ統計とは

与えられた事前分布を観測データによって

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{\theta}) \cdot \frac{p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})}$$

と修正することで事後分布$p(\boldsymbol{\theta}|\boldsymbol{y})$を求め，これに基づいてより精度の高いシステムモデルの統計的推定を行う方法

5.3 ベイズ統計

ベイズ統計におけるベイズの定理の解釈

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{\theta}) \cdot \frac{p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})}$$

$p(\boldsymbol{\theta})$	事前分布(a *priori* distribution)
$p(\boldsymbol{\theta}\|\boldsymbol{y})$	事後分布(a *posteriori* distribution)
$p(\boldsymbol{y}\|\boldsymbol{\theta})$	観測値の尤度(likelihood)
$\frac{p(\boldsymbol{y}\|\boldsymbol{\theta})}{p(\boldsymbol{y})}$	観測データによる修正項

通常は分母の$p(\boldsymbol{y})$を計算せずに　$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})$　とする.

5.3 ベイズ統計

ベイズ統計の解釈

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y})}_{\text{事後分布}} \propto \underbrace{p(\boldsymbol{\theta})}_{\text{事前分布}}\underbrace{p(\boldsymbol{y}|\boldsymbol{\theta})}_{\text{尤度（観測データ情報）}}$$

ベイズ統計の視点 → ある問題について，観測データに基づいて事前情報を修正し，より良い事後情報を得ようとしている．

モデリングの視点 → 事前情報は対象の第一原理モデルに相当する．

ベイズ統計は，観測データの収集によってモデリングの精度を向上させるという，グレーボックスモデリングの考え方に対応する．

5.3 ベイズ統計

全確率の定理

連続系の場合
$$p(x) = \int p(x|y)p(y)\mathrm{d}y$$

離散系の場合
$$p(x) = \sum_{y} p(x|y)p(y)$$

ベイズの定理（連続と離散）

連続系の場合
$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(x|y)p(x)}{\int p(y|x')p(x')\mathrm{d}x}$$

離散系の場合
$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(x|y)p(x)}{\sum_{x'} p(y|x')p(x')}$$

5.4 推定理論

ベイズ統計の統計的推測

$$\widehat{\boldsymbol{\theta}}^* = \arg\min_{\widehat{\theta}} R_d^{\widehat{\theta}}$$:ベイズ推定値

$\widehat{\boldsymbol{\theta}} = g(\boldsymbol{y})$	パラメータ推定値を与える推定則　$g(\cdot)$は非線形関数
$\boldsymbol{e} \triangleq \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}$	推定誤差ベクトル
$L_d(\boldsymbol{e}) \triangleq \Vert\boldsymbol{e}\Vert^d$ $(d = 1,2,\ldots,\infty)$	ノルムにより定義された評価関数
$R_d^{\widehat{\theta}}$	データ取得後の推定誤差の期待値(**ベイズリスク**) $\triangleq \mathrm{E}[L_d(\boldsymbol{e})\vert\boldsymbol{y}] = \int_{-\infty}^{\infty} L_d(\boldsymbol{e})p(\boldsymbol{\theta}\vert\boldsymbol{y})\mathrm{d}\boldsymbol{\theta}$

確率変数xの期待値の定義

$$\bar{x} \triangleq \mathrm{E}[x] \triangleq \int_{-\infty}^{\infty} xp(x)\mathrm{d}x$$

5.4 推定理論

ベイズリスクの計算には以下の積分が必要.

$$
\begin{aligned}
R_d^{\hat{\theta}} = \mathrm{E}[g(\boldsymbol{\theta})|\boldsymbol{y}] &= \int_{-\infty}^{\infty} g(\boldsymbol{\theta})p(g(\boldsymbol{\theta})|\boldsymbol{y})\mathrm{d}\boldsymbol{\theta} \\
&= \int_{-\infty}^{\infty} g(\boldsymbol{\theta})\frac{p\big(g(\boldsymbol{\theta})\big)p\big(\boldsymbol{y}|g(\boldsymbol{\theta})\big)}{p(\boldsymbol{y})}\mathrm{d}\boldsymbol{\theta}
\end{aligned}
$$

この計算を解析的に行うことは難しいため, 確率変数の正規分布を仮定する手法がよく用いられる.

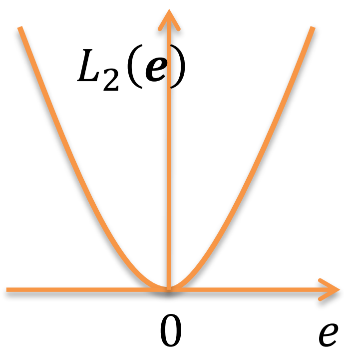
⇩

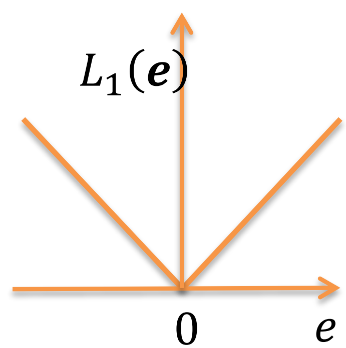
実際には正規性が仮定できない場合が多く, モンテカルロ積分を計算機で行う.
（UKFやパーティクルフィルタ）

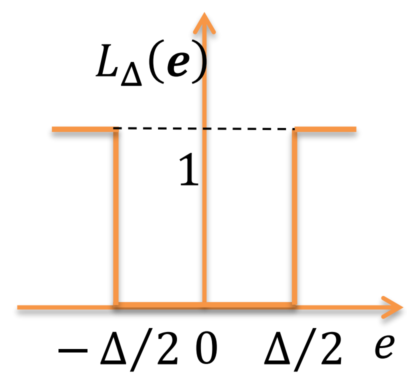
5.4 推定理論

評価関数は様々な形が考えられ，推定問題では重要である.

	二乗誤差評価	絶対誤差評価	一様誤差評価
評価関数	$L_2(\boldsymbol{e}) = \|\boldsymbol{e}\|^2$	$L_1(\boldsymbol{e}) = \|\boldsymbol{e}\|$	$L_\Delta(\boldsymbol{e}) = \begin{cases} 0 & (\|\boldsymbol{e}\| \leq \Delta/2) \\ 1 & (\|\boldsymbol{e}\| > \Delta/2) \end{cases}$
ベイズ推定値	$\hat{\boldsymbol{\theta}}^* = \mathrm{E}[\boldsymbol{\theta}\|\boldsymbol{y}]$	$\hat{\boldsymbol{\theta}}^* = \boldsymbol{\theta}_{\mathrm{median}}$	$\hat{\boldsymbol{\theta}}^* = \boldsymbol{\theta}_{\mathrm{mode}} \max_{\theta} p(\boldsymbol{\theta}\|\boldsymbol{y})$







二乗誤差評価：最小二乗推定法およびカルマンフィルタに対応.

絶対誤差評価：ロバスト推定の枠組みで用いられる.

一様誤差評価：0-1損失関数とも呼ばれる.

5.5 正規分布の場合の最尤推定法（スカラの場合）

$$y = cx + w$$

⇨ 確率変数xの推定問題について考える.

定義

$\mathrm{E}[x] = \bar{x}$　　　$\mathrm{E}[w] = \bar{w}$

$\mathrm{E}[(x-\bar{x})^2] = \sigma_x^2$　　　$\mathrm{E}[(w-\bar{w})^2] = \sigma_w^2$

信号と雑音は無相関と仮定

観測値yも確率変数

$$\bar{y} = \mathrm{E}[y] = c\bar{x} + \bar{w}$$

$$\begin{aligned}\sigma_y^2 &= \mathrm{E}[(y-\bar{y})^2] \\ &= \mathrm{E}[\{c(x-\bar{x}) + (w-\bar{w})\}^2] \\ &= c^2\sigma_x^2 + \sigma_w^2\end{aligned}$$

仮定：期待値演算の線形性と信号と雑音は無相関

5.5 正規分布の場合の最尤推定法(スカラの場合)

確率密度関数の定義

$$p_1(\xi) = \frac{\mathrm{d}}{\mathrm{d}\xi} P(x \le \xi) \qquad p_2(\eta) = \frac{\mathrm{d}}{\mathrm{d}\eta} P(w \le \eta)$$

$P(x \le \xi)$, $P(w \le \eta)$はそれぞれxとwの累計分布関数

例えば)　x_1の確率密度関数を$p_1(x_1)$

x_2の確率密度関数を$p_2(x_2)$

x_1とx_2は独立

$z = x_1 + x_2$

確率変数zの確率密度関数

$$p_3(z) = \int_{-\infty}^{\infty} p_1(x_1) p_2(z - x_1) \mathrm{d}\, x_1$$

5.5 正規分布の場合の最尤推定法（スカラの場合）

観測値yの確率密度関数

$$p_3(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} P(y \leq \theta) = \int_{-\infty}^{\infty} p_1(c\xi) p_2(\theta - c\xi) \mathrm{d}\,\xi \qquad (5.31)$$

$$\hookrightarrow p_3\,(y) = \underline{\int_{-\infty}^{\infty} p_1(cx) p_2(y - cx) \mathrm{d}\,x} \qquad (5.32)$$

この積分はどのように行う？

信号xと雑音wがともに正規性であると仮定すると，それらの確率密度関数は

$$p_1(x) = \frac{1}{\sqrt{2\pi{\sigma_x}^2}} \exp\left\{-\frac{(cx - c\overline{x})^2}{2{\sigma_x}^2}\right\} \qquad (5.33)$$

$$p_2(w) = \frac{1}{\sqrt{2\pi{\sigma_w}^2}} \exp\left\{-\frac{(w - \overline{w})^2}{2{\sigma_w}^2}\right\} \qquad (5.34)$$

となり，式(5.32)の積分を計算することが可能

5.5 正規分布の場合の最尤推定法(スカラの場合)

信号と雑音の線形変形である観測値も正規性になるので
出力yの確率密度関数は,

$$p_3(w) = \frac{1}{\sqrt{2\pi{\sigma_y}^2}} \exp\left\{-\frac{(y-\overline{y})^2}{2{\sigma_y}^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi(c^2{\sigma_x}^2+{\sigma_w}^2)}} \exp\left[-\frac{\{y-(c\overline{x}+\overline{w})\}^2}{2(c^2{\sigma_x}^2+{\sigma_w}^2)}\right] \tag{5.35}$$

で与えられる.

5.5 正規分布の場合の最尤推定法(スカラの場合)

Point5.3　正規性は線形変換で保存される

多変数確率変数$\boldsymbol{x} = [x_1,\ x_2,\ \cdots x_n]^T$の確率密度関数は

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \mathrm{det}\boldsymbol{P}_x}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\overline{\boldsymbol{x}})^T \boldsymbol{P}_x^{-1}(\boldsymbol{x}-\overline{\boldsymbol{x}})\right] \tag{5.36}$$

で与えられる. このベクトル値確率変数$\boldsymbol{x}$を

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \tag{5.37}$$

のように線形変換(厳密にはアフィン変換)して$y = [y_1,\ y_2,\ \cdots y_n]^T$を得る.
このとき, $\boldsymbol{y}$も正規分布に従い, その平均値ベクトル$\overline{\boldsymbol{y}}$と共分散行列$\boldsymbol{P}_y$は,
それぞれ次のように与えられる.

$$\overline{\boldsymbol{y}} = \boldsymbol{A}\overline{\boldsymbol{x}} + \boldsymbol{b},\ \boldsymbol{P}_y = \boldsymbol{A}\boldsymbol{P}_x\boldsymbol{A}^T \tag{5.38}$$

このように, 線形変換によって確率密度関数

$$p(\boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^n \mathrm{det}\boldsymbol{P}_y}} \exp\left[-\frac{1}{2}(\boldsymbol{y}-\overline{\boldsymbol{y}})^T \boldsymbol{P}_y^{-1}(\boldsymbol{y}-\overline{\boldsymbol{y}})\right] \tag{5.39}$$

の形は正規分布のままで保存される.

5.5 正規分布の場合の最尤推定法（スカラの場合）

ベイズの定理　$p(x|y) = \frac{p_1(x)p_2(y|x)}{p_3(y)}$　(5.40)

を利用するために，信号xが与えられたときの観測値yの
条件付確率密度関数を計算すると，

$$p_2(y|x) = \frac{1}{\sqrt{2\pi{\sigma_w}^2}}\exp\left\{-\frac{(y-cx-\overline{w})^2}{2{\sigma_w}^2}\right\} \tag{5.41}$$

となる．式(5.40)に式(5.33)，(5.35)，(5.41)を代入し計算する．
まず，式(5.40)右辺の分子の指数部の計算を行う．

$$\begin{aligned}
&\sigma_x^{-2}(x-\overline{x})^2 + \sigma_w^{-2}(y-cx-\overline{w})^2 \\
&\quad = \sigma_x^{-2}(x-\overline{x})^2 + {\sigma_w}^{-2}\{(y-\overline{y}) - c(x-\overline{x})\}^2 \\
&\quad = (\sigma_x^{-2} + c^2\sigma_w^{-2})(x-\overline{x})^2 - 2c\sigma_w^{-2}(x-\overline{x})(y-\overline{y}) + \sigma_w^{-2}(y-\overline{y})^2
\end{aligned} \tag{5.42}$$

が得られる．ただし，$-1/2$の部分は省略した．

5.5 正規分布の場合の最尤推定法(スカラの場合)

第4章の式(4.10)より,

$$\sigma_x^{-2} + c^2\sigma_w^{-2} = \sigma^{-2} \tag{5.43}$$

なので，式(5.42)は

$$\sigma^{-2}(x-\overline{x})^2 - 2c\sigma_w^{-2}(x-\overline{x})(y-\overline{y}) + \sigma_w^{-2}(y-\overline{y})^2 \tag{5.44}$$

となる．この式を平方完成すると,

$$\sigma^{-2}\{(x-\overline{x}) - c\sigma^2\sigma_w^{-2}(y-\overline{y})\}^2 + (\sigma_w^{-2} - c^2\sigma^2\sigma_w^{-4})(y-\overline{y})^2 \tag{5.45}$$

第4章の式(4.14)で与えた最小二乗推定法

$$\hat{x} = \overline{x} + c\sigma^2\sigma_w^{-2}(y-\overline{y}) \tag{5.46}$$

を用いると，式(5.45)は,

$$\sigma^{-2}(x-\hat{x})^2 + (\sigma_w^{-2} - c^2\sigma^2\sigma_w^{-4})(y-\overline{y})^2 \tag{5.47}$$

となる.

5.5 正規分布の場合の最尤推定法(スカラの場合)

式(5.47)の第2項に含まれる$\sigma_w^{-2} - c^2\sigma^2\sigma_w^{-4}$の逆数を逆行列補題を用いて計算すると,

$$\begin{aligned}(\sigma_w^{-2} - c^2\sigma^2\sigma_w^{-4})^{-1} &= \sigma_w^2 + c^2(\sigma^{-2} - c^2\sigma_w^{-2})^{-1} \\ &= \sigma_w^2 + c^2\sigma_x^2\end{aligned} \tag{5.48}$$

となる. 式(5.48)を式(5.47)に代入すると,

$$\sigma^{-2}(x-\hat{x})^2 + (c^2{\sigma_x}^2 + {\sigma_w}^2)^{-1}(y-\hat{y})^2 \tag{5.49}$$

となる. これがベイズの定理右辺の分子の指数部である. 次に, 分母の$p_3(y)$の指数部

$${\sigma_y}^{-2}(y-\overline{y})^2 = (c^2\sigma_x^2 + \sigma_w^2)(y-\overline{y})^2 \tag{5.50}$$

を考慮すると, 式(5.40)右辺全体の指数部は,

$$-\frac{(x-\hat{x})^2}{2\sigma^2} \tag{5.51}$$

となる.

5.5 正規分布の場合の最尤推定法(スカラの場合)

式(5.40)右辺の係数部を計算すると

$$\frac{\frac{1}{\sqrt{2\pi\sigma_x^2}}\frac{1}{\sqrt{2\pi\sigma_w^2}}}{\frac{1}{\sqrt{2\pi\sigma_y^2}}} = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sigma_y^2}{\sigma_x^2\sigma_w^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \tag{5.52}$$

となり，式(5.51)，(5.52)より，

$$p(x|y) = \frac{p_1(x)p_2(y|x)}{p_3(y)} = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\hat{x})^2}{2\sigma^2}\right\} \tag{5.53}$$

が得られる.

事後確率密度関数

$$p(x|y) = \frac{p_1(x)p_2(y|x)}{p_3(y)}$$

を最大にする$x = \hat{x}$を推定値とする方法を，最尤推定法という.

5.5 正規分布の場合の最尤推定法（スカラの場合）

Point5.5　最尤推定法
(ガウス＝マルコフの定理)

信号xと雑音wがともに正規性であれば，観測値yが与えられた時のxの事後確率密度関数は，

$$p(x|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\hat{x})^2}{2\sigma^2}\right\} \tag{5.54}$$

となり，最小二乗推定値

$$\hat{x} = \overline{x} + \frac{c\sigma^2}{\sigma_w^2}\{y - (c\overline{x} + \overline{w})\}$$

は最尤値に一致する．ただし，

$$\sigma^2 = \frac{1}{\sigma_x^{-2} + c^2\sigma_w^{-2}}$$