確率と最適化

—事後分布に基づく行動の決定—

今回学ぶこと

- 連続確率変数における確率推論
- 事後分布に基づく行動の決定

演習問題より

同時分布が 2 次元ガウス分布

$$P_{XY}(x,y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)$$

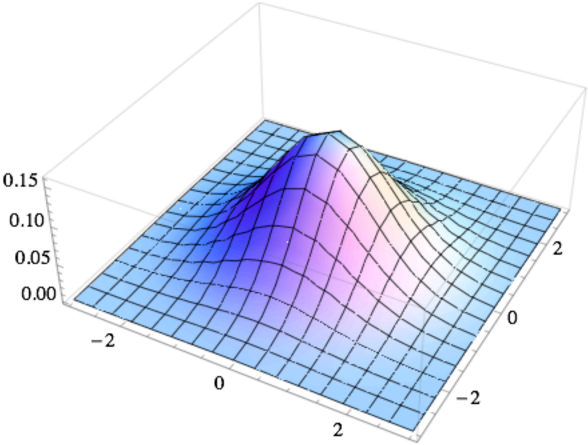
で与えられる。Y の実現値 $y^* = 2$ が観測されたとき、事後確率密度関数 $p_{X|Y}(x|y^*)$ を計算せよ。

問題の構造

- 2 変数 X, Y が連続確率変数である
- 「系に関する知識」として、同時分布を我々は知っている（系の「確率モデル」）
- Y が観測変数であり、X が推論の対象とする変数

同時分布

$$P_{XY}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right)$$



事後確率密度関数の計算 (1)

事後確率密度関数は

$$P_{X|Y}(x|y^*) = \frac{P_{XY}(x,y^*)}{P_Y(y^*)} = \frac{p_{XY}(x,y^*)}{\int_{-\infty}^{\infty} P_{XY}(x,y^*)dx}$$

であるので、まず分母を計算すると

$$\begin{aligned} \int_{-\infty}^{\infty} P_{XY}(x,2)dx &= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{x^2+2^2}{2}\right) dx && (1) \\ &= \frac{1}{e^2\sqrt{2\pi}} && (2) \end{aligned}$$

を得る。

事後確率密度関数の計算 (2)

この値を利用すると

$$P_{X|Y}(x|2) = \frac{e^2}{\sqrt{2\pi}} \exp\left(-\frac{x^2+2^2}{2}\right) \tag{3}$$

を得る。さらに変形すると

$$P_{X|Y}(x|2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{4}$$

となり、平均 0、分散 1 のガウス分布となる。ちなみに周辺分布 $P_X(x)$ も同じガウス分布となる。この問題では、X, Y が独立であるため、Y の観測値は X の推測の役には立たない。

演習問題より

事前確率密度関数が一様分布

$$P_M(m) = \begin{cases} 1/(2\alpha) & -\alpha \leq m \leq \alpha \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

条件付確率密度関数は次のガウス分布

$$P_{X|M}(x|m) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2}\right) \tag{6}$$

を仮定する。$\alpha = 5$ で $x^* = 2.3$ を観測したときの M に関する事後分布の概形を描け。

<u>問題の構造</u>

- 2 変数 X, M が連続確率変数である
- 「系に関する知識」として、事前分布と条件付き確率密度関数を我々は知っている

事後確率密度関数の計算 (1)

M に関する事後確率密度関数は

$$P_{M|X}(m|x^*) = \frac{P_{X|M}(x^*|m)P_M(m)}{P_X(x^*)} = \frac{P_{X|M}(x^*|m)P_M(m)}{\int_{-\infty}^{\infty} P_{X|M}(x^*|m)P_M(m)dm} \tag{7}$$

と与えられる。

事後確率密度関数の計算 (2)

分母の正規化定数 (エビデンス) の計算を行い、

$$\begin{aligned}\int_{-\infty}^{\infty} P_{X|M}(2.3|m)P_M(m)dm &= \int_{-5}^{5} \frac{1}{10\sqrt{2\pi}} \exp\left(-\frac{(2.3-m)^2}{2}\right) dm \\ &= \alpha (= 0.0996533 = 1/10.0348) \end{aligned} \tag{8}$$

と置く。(概形を書くためには、α の正確な値を知る必要はないことに注意)。

事後確率密度関数の計算 (3)

事後確率密度関数は

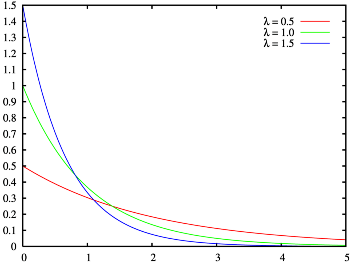
$$P_{M|X}(m|2.3) = \begin{cases} \frac{1}{10\alpha\sqrt{2\pi}} \exp\left(-\frac{(2.3-m)^2}{2}\right) & -5 \leq m \leq 5 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

となる→概形は板書にて。

指数分布

$$P_{X|L}(x|\lambda) = \begin{cases} \lambda \exp(-\lambda x), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

まれに生じる事象の生起間隔をモデル化するために、しばしば利用される。



問題

ある製品の製造から故障までにいたる時間を指数分布でモデリングしたい：

$$P_{X|L}(x|\lambda) = \begin{cases} \lambda \exp(-\lambda x), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{10}$$

パラメータ λ の事前確率を次の通り与える：

$$P_L(\lambda) = \begin{cases} \alpha \exp\left(-\frac{(\lambda-2)^2}{2}\right), & \lambda \geq 0 \\ 0, & \lambda < 0 \end{cases} \tag{11}$$

α は規格化定数である。ある製品の故障までの期間は 4 であった。この観測によって得られるパラメータの事後確率密度関数を定数係数を除いて求めよ。

解答

この問題では、定数係数を忘れて計算してもよいので、

$$
\begin{aligned}
P_{L|X}(\lambda|4) &= \frac{P_{X|L}(4|\lambda)P_L(\lambda)}{P_X(4)} && (12)\\
&= \frac{P_{X|L}(4|\lambda)P_M(\lambda)}{\int_{-\infty}^{\infty} P_{X|L}(4|\lambda)P_L(\lambda)d\lambda} && (13)\\
&\propto P_{X|L}(4|\lambda)P_M(\lambda) && (14)\\
&\propto \lambda \exp(-4\lambda)\exp\left(\frac{-(\lambda-2)^2}{2}\right) && (15)\\
&= \lambda \exp\left(\frac{-(\lambda-2)^2-8\lambda}{2}\right) && (16)
\end{aligned}
$$

となる ($\lambda \geq 0$ の場合)。$\lambda < 0$ のときは、$P_{L|X}(\lambda|4) = 0$ となる。

ベイズアプローチ

- 興味ある系の確率的モデルを構築する（モデリング）
- 観測する
- 興味ある変数の事後確率を計算する (推論)
- 事後確率分布をもとに行動を決める (行動決定)

ベイズ推論における「行動決定」

- 損失関数 (loss function)、または利得関数 (gain function) を定義
- 損失関数 (または利得関数) の事後確率による期待値を計算
- 損失の小さくなる（または利得が大きくなる）アクションを選択

損失関数・利得関数

- X: 系の状態を表す変数
- $\mathcal{A} = \{a_1, a_2, \ldots a_n\}$: 可能な「アクション」

損失関数 $L(x|a)$ は、系の状態が $x \in D(X)$ であるときに行動 $a \in \mathcal{A}$ をとった場合の「損失」を表す。損失はなるべく小さくしたいので、行動基準としては、損失関数値を小さくする行動を選びたい。重要な点として、x は直接に観測できないことに注意したい。

なお利得関数の場合もほぼ同様であるが、利得なのでなるべく利得を大きくするように行動を選ぶことになる。

損失関数の例

腫瘍マーカ検査後の医師の判断に関する損失関数 $L(x|a)$

$x \backslash a$	1(再検査を勧める)	0(再検査を勧めない)
1(腫瘍がある)	0	100
0(腫瘍がない)	10	0

- 腫瘍マーカの結果は「確率的」である（すなわち、x は直接観測できない！）
- X(腫瘍の有無) に関する事後確率は計算できる。

期待損失

与えられた $a \in \mathcal{A}$ と観測値 y^* に対して、$L(x|a)$ の事後確率分布に関する期待値

$$E[L(X|a)] = \sum_{x \in D(X)} P_{X|Y}(x|y^*) L(x|a)$$

を a に対応する期待損失と呼ぶ。

ベイズ行動選択基準

最も期待損失の小さい行動 (または最も期待利得が大きい行動)

$$a^* = \arg\min_{a \in \mathcal{A}} E[L(X|a)]$$

を実際の行動として選択する。